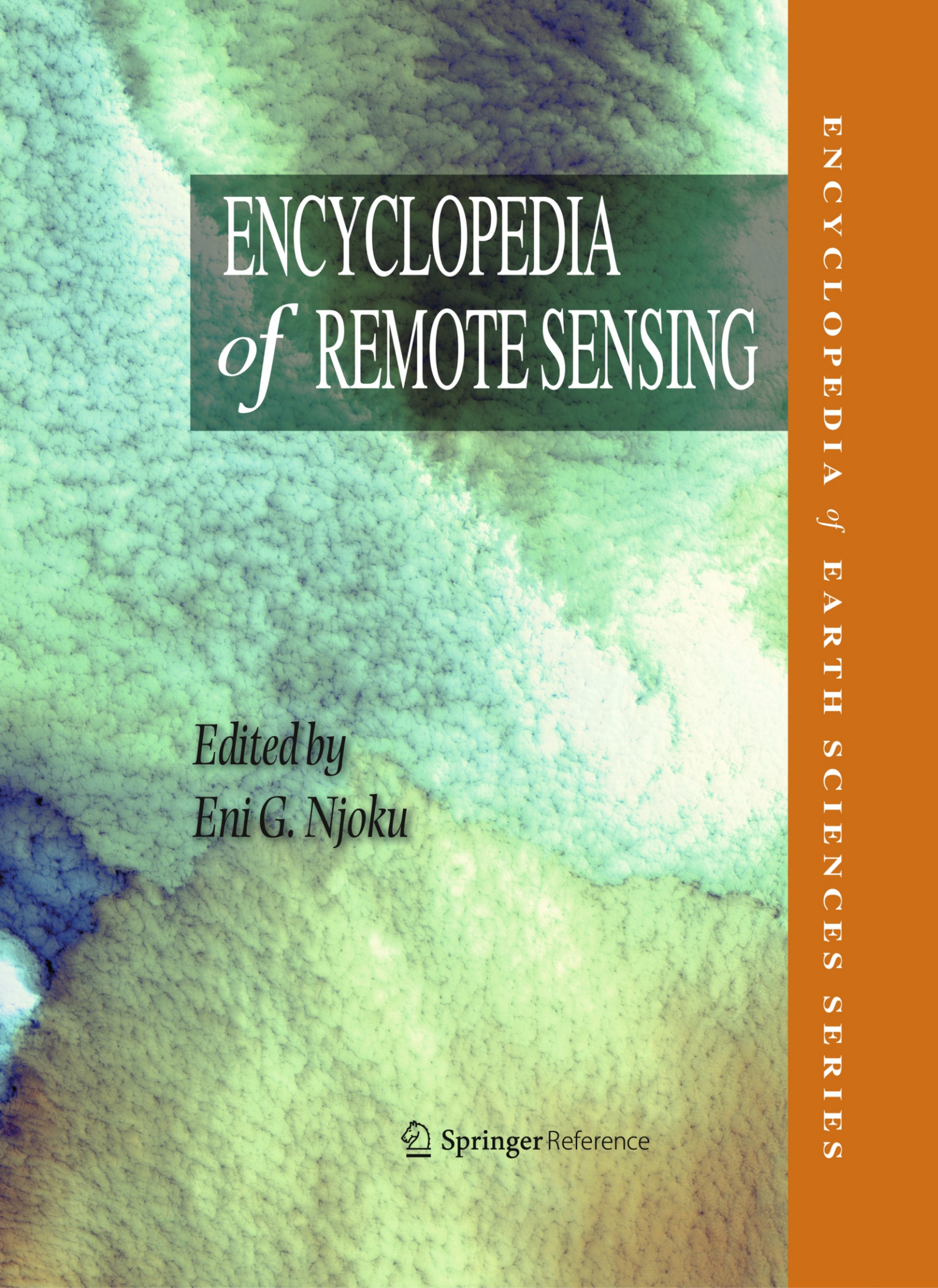




ENCYCLOPEDIA *of* REMOTE SENSING

Edited by
Eni G. Njoku



Springer Reference

ENCYCLOPEDIA *of* EARTH SCIENCES SERIES

ENCYCLOPEDIA *of* REMOTE SENSING

Encyclopedia of Earth Sciences Series

ENCYCLOPEDIA OF REMOTE SENSING

Volume Editor

Eni G. Njoku is a Senior Research Scientist at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA. He has a B.A. from the University of Cambridge, and S.M. and Ph.D. from the Massachusetts Institute of Technology. His research focuses on spaceborne microwave sensing with application to land surface hydrology and the global water cycle. Amongst his awards are the NASA Exceptional Service Medal (1985) and Fellow of the Institute of Electrical and Electronics Engineers (1995).

Section Editors

Michael J. Abrams
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA

Vincent V. Salomonson
Department of Geography
University of Utah
Salt Lake City, UT 84112
USA

Ghassem R. Asrar
World Climate Research Programme
World Meteorological Organization
1211 Geneva
Switzerland

Vernon H. Singhroy
Canada Centre for Remote Sensing
Ottawa
Ontario K1A 0Y7
Canada

Frank S. Marzano
Department of Information Engineering
Sapienza University of Rome
00184 Rome, Italy
and Centre of Excellence CETEMPS
University of L'Aquila
67100 L'Aquila
Italy

F. Joseph Turk
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA

Peter J. Minnett
Meteorology and Physical Oceanography
Rosenstiel School of Marine and Atmospheric Science
University of Miami
Miami, FL 33149
USA

Aims of the Series

The *Encyclopedia of Earth Sciences Series* provides comprehensive and authoritative coverage of all the main areas in the Earth Sciences. Each volume comprises a focused and carefully chosen collection of contributions from leading names in the subject, with copious illustrations and reference lists.

These books represent one of the world's leading resources for the Earth Sciences community. Previous volumes are being updated and new works published so that the volumes will continue to be essential reading for all professional earth scientists, geologists, geophysicists, climatologists, and oceanographers as well as for teachers and students. See the back of this volume for a current list of titles in the *Encyclopedia of Earth Sciences Series*. Go to <http://www.springerlink.com/reference-works/> and springerreference.com to visit the "Earth Sciences Series" on-line.

About the Series Editor

Professor Charles W. Finkl has edited and/or contributed to more than eight volumes in the Encyclopedia of Earth Sciences Series. For the past 25 years he has been the Executive Director of the Coastal Education & Research Foundation and Editor-in-Chief of the international *Journal of Coastal Research*. In addition to these duties, he is Professor at Florida Atlantic University in Boca Raton, Florida, USA. He is a graduate of the University of Western Australia (Perth) and previously worked for a wholly owned Australian subsidiary of the International Nickel Company of Canada (INCO). During his career, he acquired field experience in Australia; the Caribbean; South America; SW Pacific islands; southern Africa; Western Europe; and the Pacific Northwest, Midwest, and Southeast USA.

Founding Series Editor

Professor Rhodes W. Fairbridge (deceased) has edited more than 24 Encyclopedias in the Earth Sciences Series. During his career he has worked as a petroleum geologist in the Middle East, been a WW II intelligence officer in the SW Pacific and led expeditions to the Sahara, Arctic Canada, Arctic Scandinavia, Brazil and New Guinea. He was Emeritus Professor of Geology at Columbia University and was affiliated with the Goddard Institute for Space Studies.

ENCYCLOPEDIA OF EARTH SCIENCES SERIES

ENCYCLOPEDIA *of* REMOTE SENSING

edited by

ENI G. NJOKU

*Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California
USA*

 **Springer** Reference

Library of Congress Control Number: 2013953424

ISBN 978-0-387-36698-2

This publication is available also as:

Electronic publication under ISBN 978-0-387-36699-9 and

Print and electronic bundle under ISBN 978-0-387-36700-2

Springer New York, Heidelberg, Dordrecht, London

Printed on acid-free paper

Cover photo: Cloud formations over the western Aleuthian Islands, taken by Landsat 7, 1 June 2000.
Credit: US Geological Survey, Earth Resources Observation and Science (EROS) Center.

Every effort has been made to contact the copyright holders of the figures and tables which have been reproduced from other sources. Anyone who has not been properly credited is requested to contact the publishers, so that due acknowledgment may be made in subsequent editions.

All rights reserved for the contributions *Aerosols; Air Pollution; Atmospheric General Circulation Models; Calibration and Validation; Calibration, Optical/Infrared Passive Sensors; Calibration, Synthetic Aperture Radars; Cloud Properties; Data Processing, SAR Sensors; Earth System Models; Emerging Technologies; Emerging Technologies, Free-Space Optical Communications; Emerging Technologies, Radar; Emerging Technologies, Radiometer; Geodesy; Geomorphology; Geophysical Retrieval, Forward Models in Remote Sensing; Geophysical Retrieval, Inverse Problems in Remote Sensing; Geophysical Retrieval, Overview; GPS, Occultation Systems; Ionospheric Effects on the Propagation of Electromagnetic Waves; Irrigation Management; Land Surface Roughness; Land-Atmosphere Interactions, Evapotranspiration; Lidar Systems; Limb Sounding, Atmospheric; Madden-Julian Oscillation (MJO); Mission Costs of Earth-Observing Satellite; Ocean Surface Topography; Ocean-Atmosphere Water Flux and Evaporation; Precision Agriculture; Reflected Solar Radiation Sensors, Multiangle Imaging; Reflected Solar Radiation Sensors, Polarimetric; Sea Level Rise; Sea Surface Wind/Stress Vector; Solid Earth Mass Transport; Stratospheric Ozone; Terrestrial Snow; Thermal Radiation Sensors (Emitted); Trace Gases, Stratosphere, and Mesosphere; Urban Environments, Beijing Case Study; Volcanism; Water Vapor*

© Springer Science+Business Media New York 2014

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Contents

Contributors	xi	Calibration, Optical/Infrared Passive Sensors	47
Preface	xxiii	<i>Carol Bruegge</i>	
Acknowledgments	xxv	Calibration, Synthetic Aperture Radars	51
Acoustic Radiation	1	<i>Anthony Freeman</i>	
<i>Alain Weill</i>		Calibration, Scatterometers	54
Acoustic Tomography, Ocean	4	<i>David Long</i>	
<i>Brian Dushaw</i>		Climate Data Records	58
Acoustic Waves, Propagation	11	<i>Eric F. Wood</i>	
<i>Alain Weill</i>		Climate Monitoring and Prediction	58
Acoustic Waves, Scattering	13	<i>Mathew R. P. Sapiano</i>	
<i>Alain Weill</i>		Cloud Liquid Water	68
Aerosols	16	<i>Fuzhong Weng</i>	
<i>Ralph Kahn</i>		Cloud Properties	70
Agricultural Expansion and Abandonment	20	<i>Matthew Lebsock and Steve Cooper</i>	
<i>Jiaguo Qi</i>		Coastal Ecosystems	73
Agriculture and Remote Sensing	22	<i>Xiaojun Yang</i>	
<i>Jerry Hatfield and Susan Moran</i>		Commercial Remote Sensing	78
Air Pollution	32	<i>William Gail</i>	
<i>Annmarie Eldering</i>		Cosmic-Ray Hydrometeorology	83
Atmospheric General Circulation Models	35	<i>Darin Desilets and Marek Zreda</i>	
<i>Joao Teixeira, Mark Taylor, Anders Persson</i>		Cost Benefit Assessment	86
<i>and Georgios Matheou</i>		<i>Molly Macauley</i>	
Calibration and Validation	39	Crop Stress	88
<i>Andreas Colliander</i>		<i>Susan Moran</i>	
Calibration, Microwave Radiometers	46	Cryosphere and Polar Region Observing System	91
<i>Christopher Ruf</i>		<i>Mark Drinkwater</i>	

vi	CONTENTS	
Cryosphere, Climate Change Effects <i>Aixue Hu</i>	98	Emerging Technologies, Radiometer <i>Todd Gaier</i> 186
Cryosphere, Climate Change Feedbacks <i>Peter J. Minnett</i>	101	Emerging Technologies, Sensor Web <i>Mahta Moghaddam, Agnelo Silva and Mingyan Liu</i> 190
Cryosphere, Measurements and Applications <i>Roger Barry</i>	104	Environmental Treaties <i>Alexander de Sherbinin</i> 196
Data Access <i>Ron Weaver</i>	119	Fields and Radiation <i>Frank S. Marzano</i> 201
Data Archival and Distribution <i>Mark A. Parsons</i>	121	Fisheries <i>Cara Wilson</i> 202
Data Archives and Repositories <i>Ruth Duerr</i>	127	Forestry <i>Dar Roberts</i> 210
Data Assimilation <i>Dennis McLaughlin</i>	131	Gamma and X-Radiation <i>Enrico Costa and Fabio Muleri</i> 219
Data Policies <i>Ray Harris</i>	134	Geodesy <i>Calvin Klatt</i> 228
Data Processing, SAR Sensors <i>Jakob van Zyl</i>	136	Geological Mapping Using Earth's Magnetic Field <i>Vernon H. Singhroy and Mark Pilkington</i> 232
Decision Fusion, Classification of Multisource Data <i>Björn Waske and Jón Atli Benediktsson</i>	140	Geomorphology <i>David Pieri</i> 237
Earth Radiation Budget, Top-of-Atmosphere Radiation <i>Bing Lin</i>	145	Geophysical Retrieval, Forward Models in Remote Sensing <i>Eugene Ustinov</i> 241
Earth System Models <i>Andrea Donnellan</i>	146	Geophysical Retrieval, Inverse Problems in Remote Sensing <i>Eugene Ustinov</i> 247
Electromagnetic Theory and Wave Propagation <i>Yang Du</i>	150	Geophysical Retrieval, Overview <i>Eugene Ustinov</i> 251
Emerging Applications <i>William Gail</i>	159	Global Climate Observing System <i>Jean-Louis Fellous</i> 254
Emerging Technologies <i>Jason Hyon</i>	162	Global Earth Observation System of Systems (GEOSS) <i>Steffen Fritz</i> 257
Emerging Technologies, Free-Space Optical Communications <i>Hamid Hemmati</i>	163	Global Land Observing System <i>Johannes A. Dolman</i> 261
Emerging Technologies, Lidar <i>David M. Tratt</i>	177	Global Programs, Operational Systems <i>Mary Kicza</i> 263
Emerging Technologies, Radar <i>Alina Moussessian</i>	185	GPS, Occultation Systems <i>Chi O. Ao</i> 264

CONTENTS		vii
Ice Sheets and Ice Volume <i>Robert Thomas</i>	269	Microwave Dielectric Properties of Materials <i>Martti Hallikainen</i> 364
Icebergs <i>Donald L. Murphy</i>	281	Microwave Horn Antennas <i>Yahya Rahmat-Samii</i> 375
International Collaboration <i>Lisa Robock Shaffer</i>	284	Microwave Radiometers <i>Niels Skou</i> 382
Ionospheric Effects on the Propagation of Electromagnetic Waves <i>Attila Komjathy</i>	286	Microwave Radiometers, Conventional <i>Niels Skou</i> 386
Irrigation Management <i>Steven R. Evett, Paul D. Colaizzi, Susan A. O'Shaughnessy, Douglas J. Hunsaker and Robert G. Evans</i>	291	Microwave Radiometers, Correlation <i>Christopher Ruf</i> 389
		Microwave Radiometers, Interferometers <i>Manuel Martin-Neira</i> 390
Land Surface Emissivity <i>Alan Gillespie</i>	303	Microwave Radiometers, Polarimeters <i>David Kunkee</i> 395
Land Surface Roughness <i>Thomas Farr</i>	311	Microwave Subsurface Propagation and Scattering <i>Alexander Yarovoy</i> 398
Land Surface Temperature <i>Alan Gillespie</i>	314	Microwave Surface Scattering and Emission <i>David R. Lyzenga</i> 403
Land Surface Topography <i>G. Bryan Bailey</i>	320	Mission Costs of Earth-Observing Satellites <i>Randall Friedl and Stacey Boland</i> 405
Land-Atmosphere Interactions, Evapotranspiration <i>Joshua B. Fisher</i>	325	Mission Operations, Science Applications/Requirements <i>David L. Glackin</i> 407
Landslides <i>Vernon H. Singhroy</i>	328	Observational Platforms, Aircraft, and UAVs <i>Jeffrey Myers</i> 409
Law of Remote Sensing <i>Joanne Irene Gabrynowicz</i>	332	Observational Systems, Satellite <i>David L. Glackin</i> 412
Lidar Systems <i>Robert Menzies</i>	334	Ocean Applications of Interferometric SAR <i>Roland Romeiser</i> 426
Lightning <i>Rachel I. Albrecht, Daniel J. Cecil and Steven J. Goodman</i>	339	Ocean Data Telemetry <i>Michael R. Prior-Jones</i> 429
Limb Sounding, Atmospheric <i>Nathaniel Livesey</i>	344	Ocean Internal Waves <i>Werner Alpers</i> 433
Madden-Julian Oscillation (MJO) <i>Baijun Tian and Duane Waliser</i>	349	Ocean Measurements and Applications, Ocean Color <i>Samantha Lavender</i> 437
Magnetic Field <i>Nils Olsen</i>	358	Ocean Modeling and Data Assimilation <i>Detlef Stammer</i> 446
Media, Electromagnetic Characteristics <i>Yang Du</i>	362	Ocean Surface Topography <i>Lee-Lueng Fu</i> 455

viii	CONTENTS	
Ocean Surface Velocity <i>Bertrand Chapron, Johnny Johannessen and Fabrice Collard</i>	461	Radiation (Natural) Within the Earth's Environment <i>Anthony England</i> 558
Ocean, Measurements and Applications <i>Ian Robinson</i>	469	Radiation Sources (Natural) and Characteristics <i>Anthony England</i> 574
Ocean-Atmosphere Water Flux and Evaporation <i>W. Timothy Liu and Xiaosu Xie</i>	480	Radiation, Electromagnetic <i>Frank S. Marzano</i> 576
Operational Transition <i>Richard Anthes</i>	489	Radiation, Galactic, and Cosmic Background <i>David M. Le Vine</i> 581
Optical/Infrared, Atmospheric Absorption/ Transmission, and Media Spectral Properties <i>Gian Luigi Liberti</i>	492	Radiation, Multiple Scattering <i>Frank S. Marzano</i> 585
Optical/Infrared, Radiative Transfer <i>Knut Stamnes</i>	495	Radiation, Polarization, and Coherence <i>Yang Du</i> 588
Optical/Infrared, Scattering by Aerosols and Hydrometeors <i>Gian Luigi Liberti</i>	498	Radiation, Solar and Lunar <i>David M. Le Vine</i> 591
Pattern Recognition and Classification <i>Björn Waske and Jón Atli Benediktsson</i>	503	Radiation, Volume Scattering <i>Leung Tsang and Kung-Hau Ding</i> 595
Polar Ice Dynamics <i>James Maslanik</i>	509	Radiative Transfer, Solution Techniques <i>Rodolfo Guzzi</i> 606
Polar Ocean Navigation <i>Lawson Brigham</i>	512	Radiative Transfer, Theory <i>Frank S. Marzano</i> 624
Policies and Economics <i>Roberta Balstad</i>	515	Radio-Frequency Interference (RFI) in Passive Microwave Sensing <i>David Kunke</i> 634
Precision Agriculture <i>Kelly Thorp</i>	515	Rainfall <i>Ralph Ferraro</i> 640
Processing Levels <i>Ron Weaver</i>	517	Rangelands and Grazing <i>Hunt E. Raymond, Jr.</i> 653
Public-Private Partnerships <i>William Gail</i>	520	Reflected Solar Radiation Sensors, Multiangle Imaging <i>David J. Diner</i> 658
Radar, Altimeters <i>Keith Raney</i>	525	Reflected Solar Radiation Sensors, Polarimetric <i>David J. Diner</i> 663
Radar, Scatterometers <i>David Long</i>	532	Reflector Antennas <i>Yahya Rahmat-Samii</i> 668
Radar, Synthetic Aperture <i>Keith Raney</i>	536	Remote Sensing and Geologic Structure <i>Vernon H. Singhroy and Paul Lowman</i> 681
Radars <i>Keith Raney</i>	547	Remote Sensing, Historical Perspective <i>Vincent V. Salomonson</i> 684

	CONTENTS	ix
Remote Sensing, Physics and Techniques <i>David L. Glackin</i>	691	Terrestrial Snow <i>Son V. Nghiem, Dorothy K. Hall, James L. Foster and Gregory Neumann</i> 821
Resource Exploration <i>Fred A. Kruse and Sandra L. Perry</i>	702	Thermal Radiation Sensors (Emitted) <i>Simon Hook</i> 830
SAR-Based Bathymetry <i>Han Wensink and Werner Alpers</i>	719	Trace Gases, Stratosphere, and Mesosphere <i>Nathaniel Livesey</i> 834
Sea Ice Albedo <i>Donald Perovich</i>	722	Trace Gases, Troposphere - Detection from Space <i>Pietermel F. Levelt, J. P. Veeffkind and K. F. Boersma</i> 838
Sea Ice Concentration and Extent <i>Josefino C. Comiso</i>	727	Trafficability of Desert Terrains <i>Charles Hibbitts</i> 846
Sea Level Rise <i>Josh Willis</i>	743	Tropospheric Winds <i>Chris Velden</i> 849
Sea Surface Salinity <i>Gary Lagerloef</i>	747	Ultraviolet Remote Sensing <i>Arlin Krueger</i> 853
Sea Surface Temperature <i>Peter J. Minnett</i>	754	Ultraviolet Sensors <i>Arlin Krueger</i> 860
Sea Surface Wind/Stress Vector <i>W. Timothy Liu and Xiaosu Xie</i>	759	Urban Environments, Beijing Case Study <i>Son V. Nghiem, Alessandro Sorichetta, Christopher D. Elvidge, Christopher Small, Deborah Balk, Uwe Deichmann and Gregory Neumann</i> 869
Severe Storms <i>Charles A. III Doswell</i>	767	Urban Heat Island <i>Lela Prashad</i> 878
Snowfall <i>Ralf Bennartz</i>	780	Vegetation Indices <i>Alfredo Huete</i> 883
Soil Moisture <i>Yann Kerr</i>	783	Vegetation Phenology <i>John Kimball</i> 886
Soil Properties <i>Alfredo Huete</i>	788	Volcanism <i>Michael J. Abrams</i> 890
Solid Earth Mass Transport <i>Erik Ivins</i>	791	Water and Energy Cycles <i>Taikan Oki and Pat J.-F. Yeh</i> 895
Stratospheric Ozone <i>Michelle Santee</i>	796	Water Resources <i>Taikan Oki and Pat J.-F. Yeh</i> 903
Subsidence <i>Stuart Marsh and Martin Culshaw</i>	800	Water Vapor <i>Eric Fetzer</i> 909
Surface Radiative Fluxes <i>Rachel T. Pinker</i>	806	Weather Prediction <i>Peter Bauer</i> 912
Surface Truth <i>Christopher Ruf</i>	815	Wetlands <i>John Melack</i> 921
Surface Water <i>Michael Durand</i>	816	Author Index 923
		Subject Index 925

Contributors

Michael J. Abrams
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
michael.j.abrams@jpl.nasa.gov

Rachel I. Albrecht
Divisão de Satélites e Sistemas Ambientais
(DSA/CPTEC), Instituto Nacional de Pesquisas
Espaciais (INPE)
12630-000 Cachoeira Paulista, SP
Brazil
rachel.albrecht@cptec.inpe.br

Werner Alpers
Institute of Oceanography
University of Hamburg
20146 Hamburg
Germany
alpers@ifm.uni-hamburg.de

Richard Anthes
University Corporation for Atmospheric Research
Boulder, CO 80301
USA
anthes@ucar.edu

Chi O. Ao
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
chi.o.ao@jpl.nasa.gov

G. Bryan Bailey
USGS Earth Resources Observation and Science Center
Sioux Falls, SD 57198
USA
gbbailey@mchsi.com

Deborah Balk
School of Public Affairs, Baruch College
City University of New York
New York, NY
USA
deborah.balk@baruch.cuny.edu

Roberta Balstad
CIESIN
Columbia University
Palisades, NY 10964
USA
roberta@ciesin.columbia.edu

Roger Barry
National Snow and Ice Data Center
NSIDC 449 UCB
University of Colorado
Boulder, CO 80309-0449
USA
rbarry@nsidc.org

Peter Bauer
European Centre for Medium-Range Weather Forecasts
(ECMWF)
Shinfield Park
Reading RG2 9AX
UK
peter.bauer@ecmwf.int

Jón Atli Benediktsson
Faculty of Electrical and Computer Engineering
University of Iceland
107 Reykjavik
Iceland
benedikt@hi.is

Ralf Bennartz
Atmospheric and Oceanic Sciences Department
University of Wisconsin-Madison
Madison, WI 53706-1481
USA
bennartz@aos.wisc.edu

K. F. Boersma
Koninklijk Nederlands Meteorologisch Instituut (KNMI)
3732 GK, De Bilt
The Netherlands
and
Technical University Eindhoven (TUE)
5612 AZ, Eindhoven
The Netherlands
k.j.boersma@tu.nl

Stacey Boland
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA

Lawson Brigham
University of Alaska
Fairbanks, AK 99775-5840
USA
lwb48@aol.com

Carol Bruegge
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
carol.j.bruegge@jpl.nasa.gov

Daniel J. Cecil
Marshall Space Flight Center (MSFC), National
Aeronautics and Space Administration (NASA)
Huntsville, AL 35805
USA
daniel.j.cecil@nasa.gov

Bertrand Chapron
Satellite Oceanography Laboratory, IFREMER
Plouzané 29280
France
bertrand.chapron@ifremer.fr

Paul D. Colaizzi
USDA-ARS Conservation and Production Research
Laboratory
Bushland, TX 79012
USA
paul.colaiizzi@ars.usda.gov

Fabrice Collard
CLS, Division Radar
Plouzané 29280
France
dr.fab@cls.fr

Andreas Colliander
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
andreas.colliander@jpl.nasa.gov

Josefino C. Comiso
Cryospheric Sciences Laboratory, Code 615
Earth Sciences Division, NASA Goddard Space
Flight Center
Greenbelt, MD 20771
USA
josefino.c.comiso@nasa.gov

Steve Cooper
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA

Enrico Costa
Istituto di Astrofisica e Planetologia Spaziali, INAF
00133 Rome
Italy
enrico.costa@iasf-roma.inaf.it

Martin Culshaw
Honorary Research Associate, British Geological Survey,
Keyworth
Nottingham NG12 1AE
UK
and
Honorary Visiting Professor, School of Civil Engineering,
University of Birmingham, Edgbaston
Birmingham B15 2TT
UK

Alexander de Sherbinin
Center for International Earth Science Information
Network (CIESIN)
Columbia University
Palisades, NY 10964
USA
adesherbinin@ciesin.columbia.edu

Uwe Deichmann
Development Research Group, The World Bank
Washington, DC
USA
udeichmann@worldbank.org

Darin Desilets
Hydroinnova LLC
Albuquerque, NM 87106
USA
darin@hydroinnova.com

David J. Diner
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
david.j.diner@jpl.nasa.gov

Kung-Hau Ding
Air Force Research Laboratory
Wright-Patterson AFB
Dayton, OH 45433
USA

Johannes A. Dolman
Department of Earth Sciences
VU University Amsterdam
1081 Amsterdam
The Netherlands
han.dolman@vu.nl

Andrea Donnellan
Science Division
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
andrea.donnellan@jpl.nasa.gov

Charles A. III Doswell
Doswell Scientific Consulting
Norman, OK 73071
USA
cdoswell@earthlink.net

Mark Drinkwater
Mission Science Division
European Space Agency, ESA/ESTEC
2201 AZ Noordwijk ZH
The Netherlands
mark.drinkwater@esa.int

Yang Du
Zhejiang University
310027 Hangzhou
People's Republic of China
zjydu03@zju.edu.cn

Ruth Duerr
National Snow and Ice Data Center, CIRES 449 UCB,
University of Colorado
Boulder, CO 80309
USA
rduerr@nsidc.org

Michael Durand
School of Earth Sciences
The Ohio State University
275 Mendenhall Laboratory
Columbus, OH 43210
USA
durand.8@osu.edu

Brian Dushaw
Applied Physics Laboratory
University of Washington
Seattle, WA 98105-6698
USA
dushaw@apl.washington.edu

Annmarie Eldering
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
annmarie.eldering@jpl.nasa.gov

Christopher D. Elvidge
Earth Observation Group, NOAA-NESDIS National
Geophysical Data Center E/GC2
Boulder, CO
USA
chris.elvidge@noaa.gov

Anthony England
College of Engineering
University of Michigan
Ann Arbor, MI 48109
USA
england@umich.edu

Robert G. Evans
USDA-ARS
Sidney, MT 59270
USA
robert.evans@ars.usda.gov

Steven R. Evett
USDA-ARS Conservation and Production Research
Laboratory
Bushland, TX 79012
USA
steve.evett@ars.usda.gov

Thomas Farr
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
thomas.g.farr@jpl.nasa.gov

Jean-Louis Fellous
Committee on Space Research (COSPAR) Secretariat
c/o CNES-2, place Maurice Quentin
75039 Paris
France
jean-louis.fellous@cosparhq.cnes.fr

Ralph Ferraro
NOAA/NESDIS, ESSIC/CICS
College Park, MD 20740
USA
ralph.r.ferraro@noaa.gov

Eric Fetzner
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
eric.j.fetzner@jpl.nasa.gov

Joshua B. Fisher
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
joshua.b.fisher@jpl.nasa.gov

James L. Foster
Hydrological Sciences Laboratory, Code 617
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, MD
USA
james.l.foster@nasa.gov

Anthony Freeman
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
anthony.freeman@jpl.nasa.gov

Randall Friedl
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
randall.r.friedl@jpl.nasa.gov

Steffen Fritz
International Institute for Applied Systems Analysis
2361 Laxenburg
Austria
fritz@iiasa.ac.at

Lee-Lueng Fu
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
llf@jpl.nasa.gov

Joanne Irene Gabrynowicz
National Center for Remote Sensing, Air, and Space Law
The University of Mississippi School of Law
Mississippi, MS 38677-1848
USA
jgabryno@olemiss.edu

Todd Gaier
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
todd.c.gaier@jpl.nasa.gov

William Gail
Global Weather Corporation
Boulder, CO 80303
USA
wb.gail@comcast.net

Alan Gillespie
Department of Earth and Space Sciences
University of Washington
Seattle, WA 98195
USA
gillespie@ess.washington.edu

David L. Glackin
Los Angeles, CA
USA

Steven J. Goodman
National Environmental Satellite, Data, and Information
Service (NESDIS), National Oceanic and Atmospheric
Administration (NOAA)
Silver Spring, MD 20910
USA
steven.j.goodman@noaa.gov

Rodolfo Guzzi
Agenzia Spaziale Italiana ASI
00133 Roma
Italy
rodolfoguzzi@yahoo.it

Dorothy K. Hall
Cryospheric Sciences Laboratory, Code 615
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
USA
dorothy.k.hall@nasa.gov

Martti Hallikainen
Aalto University
00076 Aalto Espoo
Finland
martti.hallikainen@aalto.fi

Ray Harris
Department of Geography
University College London
London WC1E 6BT
UK
ray.harris@ucl.ac.uk

Jerry Hatfield
National Laboratory for Agriculture and the Environment
Ames, IA 50011
USA
jerry.hatfield@ars.usda.gov

Hamid Hemmati
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
hamid.hemmati@jpl.nasa.gov

Charles Hibbitts
Applied Physics Laboratory
Laurel, MD 20723
USA
karl.hibbitts@jhuapl.edu

Simon Hook
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
simon.j.hook@jpl.nasa.gov

Aixue Hu
Climate and Global Dynamics Division
National Center for Atmospheric Research
Boulder, CO 80305
USA
ahu@ucar.edu

Alfredo Huete
Plant Functional Biology and Climate Change Cluster
Faculty of Science
University of Technology
2007 Sydney, NSW
Australia
alfredo.huete@uts.edu.au

Douglas J. Hunsaker
USDA-ARS
Maricopa, AZ 85138
USA
doug.hunsaker@ars.usda.gov

E. Raymond Hunt, Jr.
USDA-ARS Hydrology and Remote Sensing Laboratory
Beltsville, MD 20705
USA
raymond.hunt@ars.usda.gov

Jason Hyon
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
jason.j.hyon@jpl.nasa.gov

Erik Ivins
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 9109
USA
erik.r.ivins@jpl.nasa.gov

Johnny Johannessen
Nansen Environmental and Remote Sensing Center
5006 Bergen
Norway
johnny.johannessen@nersc.no

Ralph Kahn
NASA Goddard Space Flight Center
Greenbelt, MD 20771
USA
ralph.kahn@nasa.gov

Yann Kerr
CNES/CESBIO
31401 Toulouse
France
yann.kerr@cesbio.cnes.fr

Mary Kicza
National Oceanic and Atmospheric Administration
(NOAA)
Washington, DC 20230
USA
nina.jackson@noaa.gov

John Kimball
Flathead Lake Biological Station
University of Montana
Polson, MT 59860-6815
USA
johnk@ntsug.umt.edu

Calvin Klatt
Geodetic Survey Division
Natural Resources Canada
Ottawa, ON K1A 0E9
Canada
cklatt@nrcan.gc.ca

Attila Komjathy
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
attila.komjathy@jpl.nasa.gov

Arlin Krueger
Atmospheric Chemistry and Dynamics Laboratory
(Code 614)
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
USA
akrueger3@verizon.net

Fred A. Kruse
Physics Department and Remote Sensing Center
Naval Postgraduate School
Monterey, CA 93943
USA
fakruse@nps.edu

David Kunkee
The Aerospace Corporation
Los Angeles, CA 90009
USA
david.b.kunkee@aero.org

Gary Lagerloef
ESR
Seattle, WA 98121
USA
lager@esr.org

Samantha Lavender
Pixalytics Ltd
Plymouth, Devon PL6 8BX
UK
slavender@pixalytics.com

David M. Le Vine
Code 615, Cryospheric Sciences Branch
NASA/Goddard Space Flight Center
Greenbelt, MD 20771
USA
david.m.levine@nasa.gov

Matthew Lebsack
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
matthew.d.lebsack@jpl.nasa.gov

Pietermel F. Levelt
Koninklijk Nederlands Meteorologisch Instituut (KNMI)
3730 AE De Bilt
The Netherlands
and
Delft University of Technology
5612 AE Eindhoven
The Netherlands
pietermel.levelt@knmi.nl

Gian Luigi Liberti
CNR/ISAC
00133 Rome
Italy
g.liberti@isac.cnr.it

Bing Lin
NASA Langley Research Center, MS 420
Hampton, VA 23681-2199
USA
bing.lin@nasa.gov

Mingyan Liu
Electrical and Computer Engineering
University of Michigan
Ann Arbor, MI 48109
USA

W. Timothy Liu
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
w.t.liu@jpl.nasa.gov

Nathaniel Livesey
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
nathaniel.j.livesey@jpl.nasa.gov

David Long
Department of Electrical and Computer Engineering
BYU Center for Remote Sensing
Brigham Young University
Provo, UT 84602
USA
long@byu.edu

Paul Lowman
NASA Goddard, Code 698.0
Greenbelt, MD 20771
USA
paul.d.lowman@nasa.gov

David R. Lyzenga
College of Engineering, Naval Architecture and Marine
Engineering
University of Michigan
Ann Arbor, MI 48109-2145
USA
lyzenga@umich.edu

Molly Macauley
Resources for the Future
Washington, DC 202-328-5043
USA
macauley@rff.org

Stuart Marsh
Nottingham Geospatial Institute
The University of Nottingham
Nottingham Geospatial Building, Triumph Road
Nottingham NG7 2TU
UK
ngi@nottingham.ac.uk

Manuel Martin-Neira
European Space Agency (ESA-ESTEC)
Keplerlaan 1
2200 Noordwijk
The Netherlands
manuel.martin-neira@esa.int

Frank S. Marzano
Department of Information Engineering
Sapienza University of Rome
00184 Rome
Italy
and
Centre of Excellence CETEMPS
University of L'Aquila
67100 L'Aquila
Italy
frank.marzano@uniroma1.it

James Maslanik
Department of Aerospace Engineering Sciences
University of Colorado, CCAR
Boulder, CO 80309
USA
jimm@colorado.edu

Georgios Matheou
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
georgios.matheou@jpl.nasa.gov

Dennis McLaughlin
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139
USA
dennism@mit.edu

John Melack
Department of Ecology, Evolution and Marine Biology
University of California
Santa Barbara, CA 93106
USA
melack@lifesci.ucsb.edu

Robert Menzies
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
robert.t.menzies@jpl.nasa.gov

Peter J. Minnett
Rosenstiel School of Marine and Atmospheric Science
University of Miami
Miami, FL 33149
USA
pminnett@rsmas.miami.edu

Mahta Moghaddam
Electrical Engineering – Electrophysics
University of Southern California
Los Angeles, CA 0089
USA
mahta@usc.edu

Susan Moran
USDA ARS Southwest Watershed Research Center
Tuscon, AZ 85719
USA
susan.moran@ars.usda.gov

Alina Moussessian
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
alina.moussessian@jpl.nasa.gov

Fabio Muleri
Istituto di Astrofisica e Planetologia Spaziali, INAF
00133 Rome
Italy
fabio.muleri@lasf-roma.inaf.it

Donald L. Murphy
International Ice Patrol, US Coast Guard
New London, CT 06320
USA
iipcomms@uscg.mil

Jeffrey Myers
NASA/Ames Research Center, Airborne Science and
Technology Laboratory
University of California, Santa Cruz
MS244-15
Moffett Field, CA 94035
USA
jmyers@mail.arc.nasa.gov

Gregory Neumann
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
gregory.neumann@jpl.nasa.gov

Son V. Nghiem
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
son.v.nghiem@jpl.nasa.gov

Taikan Oki
Institute of Industrial Science, University of Tokyo
153-8505 Tokyo
Japan
taikan@iis.u-tokyo.ac.jp

Nils Olsen
DTU Space
Technical University of Denmark
2800 Kgs. Lyngby
Denmark
nio@space.dtu.dk

Susan A. O'Shaughnessy
USDA-ARS Conservation and Production Research
Laboratory
Bushland, TX 79012
USA
susan.oshaughnessy@ars.usda.gov

Mark A. Parsons
Center for a Digital Society
Rensselaer Polytechnic Institute
Troy, NY 12180
USA
parsonsm@nsidc.org

Donald Perovich
USACE Cold Regions Research and Engineering
Laboratory
Hanover, NH 03755-1250
USA
donald.k.perovich@erdc.usace.army.mil

Sandra L. Perry
Perry Remote Sensing, LLC
Denver, CO 80231
USA
sandyp@rm.incc.net

Anders Persson
United Kingdom Meteorological Office
Exeter
Devon, EX1 3PB
UK

David Pieri
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
dave.pieri@jpl.nasa.gov

Mark Pilkington
Geological Survey of Canada
Ottawa, ON K1A 0E9
Canada
mark.pilkington@nrcan-rncan.gc.ca

Rachel T. Pinker
Department of Atmospheric and Oceanic Science
University of Maryland
College Park, MD 20742
USA
pinker@atmos.umd.edu

Lela Prashad
School of Earth and Space Exploration, 100 Cities Project
Arizona State University
Tempe, AZ 85287-1404
USA
lprashad@asu.edu

Christopher Ruf
Department of Atmospheric, Oceanic and Space Sciences
University of Michigan
Ann Arbor, MI 48109
USA
cruf@umich.edu

Michael R. Prior-Jones
British Antarctic Survey
Cambridge CB3 0ET
UK
michael@randominformation.co.uk

Vincent V. Salomonson
Department of Geography, University of Utah
South Jordan, UT 84095
USA
vincent.v.salomonson@nasa.gov

Jianguo Qi
Department of Geography/CGCEO
Michigan State University
East Lansing, MI 48823
USA
qi@msu.edu

Michelle Santee
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
michelle.l.santee@jpl.nasa.gov

Yahya Rahmat-Samii
Department of Electrical Engineering
University of California at Los Angeles
Los Angeles, CA 90095
USA
rahmat@ee.ucla.edu

Mathew R. P. Sapiano
University of Maryland - College Park
College Park, MD 20742
USA
msapiano@atmos.colostate.edu

Keith Raney
Applied Physics Laboratory
Johns Hopkins University
Laurel, MD 20723
USA
keith.raney@jhuapl.edu

Lisa Robock Shaffer
MC 0553 Rady School of Management
University of California, San Diego
La Jolla, CA 92093-0553
USA
lshaffer@ucsd.edu

Dar Roberts
Department of Geography
University of California
Santa Barbara, CA 93106
USA
dar@geog.ucsb.edu

Agnelo Silva
Electrical Engineering – Electrophysics
University of Southern California
Los Angeles, CA 0089
USA

Ian Robinson
Ocean and Earth Science
University of Southampton, at National Oceanography
Centre
Southampton SO14 3ZH
UK
isr@noc.soton.ac.uk

Vernon H. Singhroy
Applications Development Section
Natural Resources Canada
Canada Centre for Remote Sensing
Ottawa, ON K1A 0Y7
Canada
vern.singhroy@nrcan-rncan.gc.ca

Roland Romeiser
Rosenstiel School of Marine and Atmospheric Science
University of Miami
Miami, FL 33149-1031
USA
rromeiser@rsmas.miami.edu

Niels Skou
National Space Institute
Technical University of Denmark
2800 Lyngby
Denmark
ns@space.dtu.dk

Christopher Small
Lamont Doherty Earth Observatory
Marine Geology and Geophysics
Columbia University
Palisades, NY
USA
cs184@columbia.edu

Alessandro Sorichetta
Dipartimento di Scienze della Terra “A. Desio”
Universita’ degli Studi di Milano
20122 Milan
Italy
alessandro.sorichetta@unimi.it

Detlef Stammer
Institut für Meereskunde, Zentrum für Marine und
Atmosphärische Wissenschaften
Universität Hamburg
20146 Hamburg
Germany
detlef.stammer@zmaw.de

Knut Stamnes
Stevens Institute of Technology, Castle Point on Hudson
Hoboken, NJ 07030-5991
USA
kstamnes@stevens.edu

Mark Taylor
Sandia National Laboratory
Albuquerque, New Mexico 91109
USA
mataylor@sandia.gov

Joao Teixeira
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
teixeira@jpl.nasa.gov

Robert Thomas
Sigma Space
66-400 Gorzow Wlkp
Poland
robert_thomas@hotmail.com

Kelly Thorp
USDA-ARS U.S. Arid-Land Agricultural Research
Center
Maricopa, AZ 85138
USA
kelly.thorp@ars.usda.gov

Baijun Tian
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
baijun.tian@jpl.nasa.gov

David M. Tratt
The Aerospace Corporation
Los Angeles, CA 90009-2957
USA
dtratt@aero.org

Leung Tsang
Paul Allen Center
Department of Electrical Engineering
University of Washington
Seattle, WA 98195-2500
USA
leung@ee.washington.edu

Eugene Ustinov
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
eugene.a.ustinov@jpl.nasa.gov

Jakob van Zyl
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
jakob.j.vanzyl@jpl.nasa.gov

J. P. Veefkind
Koninklijk Nederlands Meteorologisch Instituut (KNMI)
3732 GK, De Bilt
The Netherlands
and
Eindhoven University of Technology
5612 AE Eindhoven
The Netherlands
veefkind@knmi.nl

Chris Velden
University of Wisconsin, CIMSS
Madison, WI 53706
USA
chrisv@ssec.wisc.edu

Duane Waliser
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 9109
USA
duane.e.waliser@jpl.nasa.gov

Björn Waske
Institute of Geodesy and Geoinformation
University of Bonn
53115 Bonn
Germany
wf@ipb.uni-bonn.de

Ron Weaver
National Snow and Ice Data Center, Cooperative Institute
for Research in Environmental Sciences
University of Colorado
Boulder, CO 80309-0449
USA
weaverr@kryos.colorado.edu

Alain Weill
Bur. Jussieu
LATMOS, Laboratoire Atmosphere Milieux
Observations Spatiales
75005 Paris
France
alain.weill@latmos.ipsl.fr

Fuzhong Weng
Center for Satellite Applications and Research (STAR)
National Oceanic and Atmospheric Administration
College Park, MD 20740
USA
fuzhong.weng@noaa.gov

Han Wensink
ARGOSS BV
8325ZH Vollenhove
The Netherlands
wensink@argoss.nl

Josh Willis
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
joshua.k.willis@jpl.nasa.gov

Cara Wilson
Southwest Fisheries Science Center
NOAA/NMFS, Environmental Research Division
Pacific Grove, CA 93950-2097
USA
cara.wilson@noaa.gov

Eric F. Wood
Department of Civil and Environmental Engineering
Princeton University
Princeton, NJ 08544
USA
efwood@princeton.edu

Xiaosu Xie
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109
USA
xiaosu.xie@jpl.nasa.gov

Xiaojun Yang
Department of Geography
Florida State University
Tallahassee, FL 32306-2190
USA
xyang@fsu.edu

Alexander Yarovoy
Delft University of Technology
2628 CN Delft
The Netherlands
a.yarovoy@tudelft.nl

Pat J.-F. Yeh
Department of Civil and Environmental Engineering
National University of Singapore
117576 Singapore
Singapore
ceeyehj@nus.edu.sg

Marek Zreda
Department of Hydrology and Water Resources
University of Arizona
Tucson, AZ 85721
USA
marek@hwr.arizona.edu

Preface

During the past few decades, the emergence of remote sensing as a discipline – its science, instruments, missions, and applications – has inspired new and comprehensive studies of the Earth. Detailed observations of Earth's land, ocean and atmospheric processes, and measurements of hitherto unexplored geophysical phenomena have been made possible by remote sensing instruments on ground-based, airborne, and spaceborne platforms. In particular, the unique vantage point of space provides spatially extensive and global perspectives of Earth. Frequent measurements, made hourly, daily, or weekly, over extended periods of years to decades, depending on the observing system and its configuration, have enabled comprehensive studies of Earth's global system. Remote sensing has thus profoundly altered our understanding of the world in which we live, and has revolutionized the approaches we use to study our environment. Each year the growing number of Earth observing satellites, and the increasingly huge amounts of data and information provided, yield new knowledge and greater appreciation of the changes occurring on our planet, with important implications for future generations of Earth inhabitants. This encyclopedia is a comprehensive reference work on Earth remote sensing that presents the foundations, principles, and state of the art of remote sensing and describes the diverse applications it serves. It covers the concepts, techniques, instrumentation, data analysis, interpretation, and applications of remote sensing. This volume is part of the Encyclopedia of Earth Science series and is organized in the same style as other volumes in the series. The scientific disciplines covered by the series have all benefited in one way or another from the new understanding and discoveries afforded by remote sensing. It is thus timely for publication of an encyclopedia that can link these disciplines and the remote sensing techniques relevant to them in an integrated framework.

The focus of the encyclopedia is on remote sensing of Earth – its atmosphere, oceans, cryosphere, and land

surface and subsurface. Some of the techniques described in this volume have their origins in the disciplines of astronomy and astrophysics, and the study of the stars and planets for which, until recently, remote sensing was the only means of obtaining observational scientific data. When applied to Earth, these techniques have blossomed into a remarkably diverse and increasingly sophisticated set of scientific, technological, and computational approaches that all fall under the umbrella of remote sensing. The rapid growth of remote sensing as a discipline is evidenced by the large number of scientific journals now devoted to this field, and the number of courses and degree programs offered at universities around the world. The measurement and interpretation of radiation scattered and emitted by Earth's atmosphere, surface, and subsurface is what we generally mean when we speak of Earth remote sensing. These measurements are obtained by instruments on remote platforms that include satellites, aircraft, balloons, drones, trucks, and stationary towers. Remote sensing instruments take many forms and are designed to measure electromagnetic radiation in specific wavelength regions of the broad electromagnetic spectrum; some instruments use other forms of radiation such as acoustic radiation. Measurements from the wide array of instruments, operating on the variety of available platforms available, can be processed and analyzed to extract characteristic information about Earth and its constituent biological, chemical, and physical structures, at resolutions from centimeters to thousands of kilometers. This remotely sensed information can be used on its own or combined with direct or 'in situ' measurements and geophysical models to give a more comprehensive understanding of the diversity of Earth science phenomena, some of which would be very limited without the unique perspective brought by remote sensing.

It is clear that an attempt to fully cover the breadth and depth of topics in remote sensing is a daunting task. Nevertheless, the need for a compendium that can be used as a reference work for this field, as a living document that

can be updated periodically to capture new advances, is a pressing one. It is with this aim in mind that the Springer *Encyclopedia of Remote Sensing* was conceived. Both this print version of the encyclopedia, which can be updated with revisions once every several years, and an online version, which can be updated on a more frequent basis by authors of individual entries, are provided. The online version can accommodate introduction of new entries as the need for new topics or treatments emerges. The encyclopedia entries cover topics that include broad introductory surveys as well as more in-depth treatment of some subjects. The entries treat topics of the physical principles of remote sensing in different wavelength regimes, propagation and scattering of radiation, geophysical models, remote sensing instrumentation, retrieval methods, remote sensing platforms and observational configurations. The models and retrieval methods are described with reference to specific applications in atmosphere, ocean, cryosphere, land, and solid earth geophysics. These applications include human impacts of climate change, and the enabling interdisciplinary science, as well as applications of direct societal benefit such as human health, food security, and prediction and mitigation of natural hazards. Earth remote sensing from space has flourished in the past few decades, and has become a truly global enterprise through development of international collaborations and partnerships, with investments from an increasing number of countries in building and operating satellite observational systems. Several entries in this volume have been devoted to describing these programs, and associated international policies and principles.

This encyclopedia is designed to support the needs of students, teachers, and professionals across a broad

spectrum of science, technology, and societal applications related to Earth remote sensing. The intended audience includes those with observational interests in the fields of oceanography, atmospheric sciences, meteorology, climate, cryospheric studies, hydrology, geology, solid earth geophysics, ecology, agronomy, forestry, environmental pollution, geography, land use and social studies, among others. The target audience also includes those with interests in remote sensing theory and practice, electromagnetic propagation, radiative transfer modeling, remote sensing instruments, spacecraft systems and orbits, environmental policy and decision-making, resource planning, and monitoring and forecasting of extreme events and natural hazards. In the commercial sector, economists, legal and insurance companies, and commercial and industrial concerns relying on the production, marketing and availability of value-added remote sensing products will also find the encyclopedia a valuable resource. The entries are presented in alphabetical order with titles that are designed to aid searches for specific topics. Cross-referencing using keywords to related entries is also provided to support efficient searches for information of interest to readers. The entries provide bibliographies for further in-depth reading. In summary, though it cannot be claimed that this encyclopedia represents an exhaustive treatment or complete coverage of the field of Earth remote sensing, it is hoped that the volume will serve as a comprehensive and dynamic introduction, and initial entry point, to inspire further reading and study of this exciting and rapidly developing field.

September 2013

Eni G. Njoku

Acknowledgments

A work the size of this encyclopedia inevitably relies on the help and cooperation of a large number of people, only some of whom can be individually identified and thanked here. My particular thanks go to the Board of Section Editors, a group of diverse and highly respected remote sensing scientists. To Mike Abrams, Ghassem Asrar, Frank Marzano, Peter Minnett, Vince Salomonson, Vern Singhroy, and Joe Turk, thank you for keeping this project on course by helping to choose the topics that form the entries, suggesting high-quality authors, reviewing the initial manuscripts, and finally checking proofs with your selected groups of authors as well as writing important contributions yourselves. I also wish to acknowledge the great help of Roberta Balstad, Farouk El-Baz, Moustafa Chahine, Jean-Marie Dubois, A.J. Chen, Robert Gurney, Jim Smith, and Guido Visconti who contributed to the early selection of topics and authors for the encyclopedia. My thanks are due also to Tom Farr who assisted me with the editorial duties during a critical stage in the project.

This leads to the largest group I wish to acknowledge, the authors of the 170 entries that range in size from several hundred words up to major contributions of several thousand words. Many authors took on more than one entry within their specialty area. Due to the length of time required to produce a volume of this type many

authors who submitted entries early had to wait a considerable amount of time before their entries were finally published. To these authors I especially wish to express my thanks for their patience and dedication to the completion of the project, and I hope they find the final volume worth the wait.

At the production end of the project has been the staff of Springer. Their help, understanding, and cooperation, especially when problems needed to be overcome, is something that cannot be appreciated enough. Their patient discussions and exchanges with the authors and board members did much to maintain the smooth progress of the project. Special acknowledgment should go to Petra van Steenberg and Sylvia Blago who provided me with encouragement through many difficult periods, and to Simone Giesler, all of who were involved with the encyclopedia from start to finish. I must also acknowledge the rewarding environment of my host institution and colleagues, all of whom provided a rich source of motivation and ideas in the field of remote sensing to inspire a publication of this type. My final appreciation goes to my wife Mary whose patience and support over the years has been a major factor in my ability to undertake this task, and to my son Eni Jr. who reminds me constantly of the power of positive thinking.

A

ACOUSTIC RADIATION

Alain Weill

Bur. Jussieu, LATMOS, Laboratoire Atmosphere Milieux
Observations Spatiales, Paris, France

Definition

Acoustic. One branch of physics which studies sound.

The word acoustic comes from the Greek word *akoustikos*. “which is related to hearing.”

Sound. It comes from the Latin word *sonum*: “which is related to the hearing sensation created by perturbation of the material medium (elastic, fluid, solid).” In physics, it is a vibration, generally in a gas, created by expansion and compression of gas molecules. Sound waves propagate in the fluid medium and do not propagate in the vacuum. Sounds can be produced in the atmosphere and oceans by living animals or by structures through interaction with the wind, as, for example, trees murmuring, mountains roaring, river sounds, and waves breaking and can be created by various instruments such as music instruments, microphones, speakers, and transducers and also by instruments developed for remote sensing such as SONAR (Sound Navigation and Ranging), ADCP (Acoustic Doppler Current Profiler), and SODAR (Sound Detection and Ranging or what are called echo sounders for atmosphere and ocean). A sound propagating in a medium is characterized by its speed c :

$$c^2 = \partial P / \partial \rho \quad (1)$$

where P is the pressure and ρ the density, and ∂ is derived.

In a gas,

$$c = (\gamma P / \rho)^{1/2} \quad (2)$$

where γ is the heat capacity ratio.

Notice that sound speed in the air for standard conditions of temperature and pressure near the surface is close to 340 m/s, while at the ocean surface it is close to 1,500 m/s, which is faster. This will have an incidence on different ways for acoustic signal processing to be done in the ocean and atmosphere.

Sound or rather a sound wave is a mechanical pressure oscillation, which is generally longitudinally propagating. *Period* T . It is the signal duration corresponding to the time when the sound wave is reproduced identically. *Frequency*. $f = 1/T$ (T in s and f in hertz). *Frequency audio spectrum* (distribution of acoustic energy as function of frequencies can be divided in four zones related to human hearing power: 0–20 Hz infrasound (not audible), 20–300 Hz is low-pitched, 300–6,000 Hz is medium range, 6,000–20,000 Hz is high-pitched, more than 20,000 Hz are ultrasonic sounds (not audible).

Sound amplitude. It corresponds to acoustic pressure fluctuation of the medium Δp (amount of energy in the sound wave) measured at one point of a surface S . It is the ratio of pressure P by the surface element S .

$$I = P / S \text{ in } \text{W/m}^2$$

For a spherical acoustic source, the intensity at distance r is

$$I r = P \text{ (sound power of the source)} / 4\pi r^2$$

Radiation. It is the way acoustic wave energy radiates and concerns acoustic rays from the acoustic sources through

the concerned medium. For example, from a microphone radiating along different directions, we are interested in the radiation diagram corresponding to the knowledge of *rays* (analogy with optical rays) along different directions.

Introduction

We shall begin to analyze (sound) or (acoustic) waves and the wave equation from which we are able to describe energy propagation and rays. If one compares them to electromagnetic waves, acoustic waves are simpler and can be described by the velocity potential, which is a scalar. We shall apply the principles of acoustic wave radiation to different types of acoustic sources used such as monopolar, dipolar sources and the response of the medium at different distances showing that the acoustic field will change of as the characteristics do so. These considerations are really fundamental for remote sensing with acoustic sounders and they are similar to electromagnetic waves, though propagation equations are different. They give information about distances when field characteristics will present some useful organization properties.

Wave equation and acoustic waves in flows

It corresponds to the study of a *plane wave* transmitted in a flow by a vibrating plane. It can be, for example, the diaphragm of a microphone vibrating along an axis x .

A *plane acoustic wave* is a general concept from the physics of waves. It corresponds to a wave where *wave fronts* (surfaces of the same phase) are infinite planes, perpendicular to the same direction of propagation.

The equation of pressure variation p (or wave propagation equation) is

$$\partial^2 p / \partial x^2 = [1/c^2] \partial^2 p / \partial t^2 \quad (3)$$

where, as already said, c is the sound speed.

For a sinusoidal wave, the solution of the equation is

$$p(x, t) = p_0 \cos[(2\pi/T)(t - x/c)] \quad (4)$$

with the wave number $k = 2\pi/T_c = 2\pi/\lambda$ and λ is the wave length.

We get

$$p(x, t) = p_0 \cos(2\pi t/T - kx) \quad (5)$$

where $2\pi/T$ is the *pulsation* ω .

Generally, one uses the complex notation:

$$p(x, t) = p_0 \exp(j\omega t - kx) \quad (6)$$

A linear relation between the displacement gradient (compressibility) gives $p = -\kappa \partial\zeta/\partial x$ (where κ is a coefficient of compressibility) from which we have, solving the sound equation:

$$\partial\zeta/\partial t = (1/(\rho_0 c)) \exp(j(\omega t - kx)) \quad (7)$$

This means that particle velocity is in phase with acoustic pressure.

Other derived definitions are useful parameters for remote sensing techniques as the acoustic impedance Z , which is the ratio between pressure and the complex amplitude of the particle.

To illustrate the interest in using acoustic impedance, in a project for the Titan satellite sounding (Weill and Blanc, 1987), it was suggested to use acoustic impedance from the satellite's surface to discriminate, by acoustic remote sensing, between solid and liquid surface just before a possible crash of the rocket at the satellite's surface.

For the flow, it is equal to $Z_c = \rho_0 c$.

For the plane wave (or *progressive wave*) $\forall x$, we have $Z(x) = Z_c$.

The *condensation* of the wave is the spatial derivative $-\partial\zeta/\partial x$, which corresponds to a relative change of density.

Monopoles, dipoles, and pulsing sphere

If wave propagation Eq. 3 is satisfied, we can work with harmonic solutions of the equation and use wave superposition in the Fourier space. Let us consider one source S of strength q radiating at radial distance r and the solution of the equation for particles radial velocities is

$$v = A (\exp(-jkr)/r^2 + jk \exp(-jkr))/r \quad (8)$$

A is a constant and the boundary conditions are such that the solution vanishes at infinity. It is important to notice that radiation behavior is different as function of distance.

The acoustic flux Fa of the radial velocity v across a sphere of radius is $id v^* 4\pi r^2$ (sphere considered at the distance r) is

$$Fa = 4\pi A (\exp(-jkr) + jkr \exp(-jkr)) \quad (9)$$

Therefore, when kr is small, the first term of (Eq. 9) predominates and the conditions correspond to what is called the *near field* (velocity in phase with the source), and when kr is large, we are in the *far field* conditions (velocity 90° in advance with the source). These very simple statements are very important in acoustic remote sensing, if (for example) active sea foam, which is an acoustic transmitter at the sea surface, has to be modeled; see Vagle and Farmer (1992) to understand acoustic noise below the surface and bubble sound emission.

A more general representation of acoustic sources corresponds to dipolar sources constituting two radiating sources of strengths or magnitudes $q+$ and $q-$ separated by a distance a and such that $m = qa$ is the *dipolar momentum* (as considered in electromagnetism).

Solving the wave equation for the two monopoles with θ the angle between r and the direction of the dipole gives two velocity components: (a) one for the near field (small r): $m(1 + jkr) \exp(-jkr) \cos(\theta)/4\pi r^3$, to which is added a transverse component varying as $\sin \theta$ (orthogonal component), and (b) one for the far field (large r): $-k^2 m \cos(\theta) \exp(-jkr)/4\pi r$, which is typically a radial component. Dipolar sources or a combination of dipolar sources are very useful to simulate acoustic antennas and



Acoustic Radiation, Figure 1 Realization of a three offset antennas for a 6 kHz minisounder. One antenna is *vertical* and two others are *slanting*. One distinguishes compression chambers as acoustic sources and the horns and antennas parts (*parabolic portion* + *circular aperture covered by acoustic foam*).

acoustic sources. Moreover, field analysis considering the distance where the field can be considered as far is very important to interpret, for example, signals coming from *active systems* (which transmit acoustic waves) or *passive systems* (which only receive acoustic waves as *acoustic radiometers* by analogy with electromagnetic and optical radiometers).

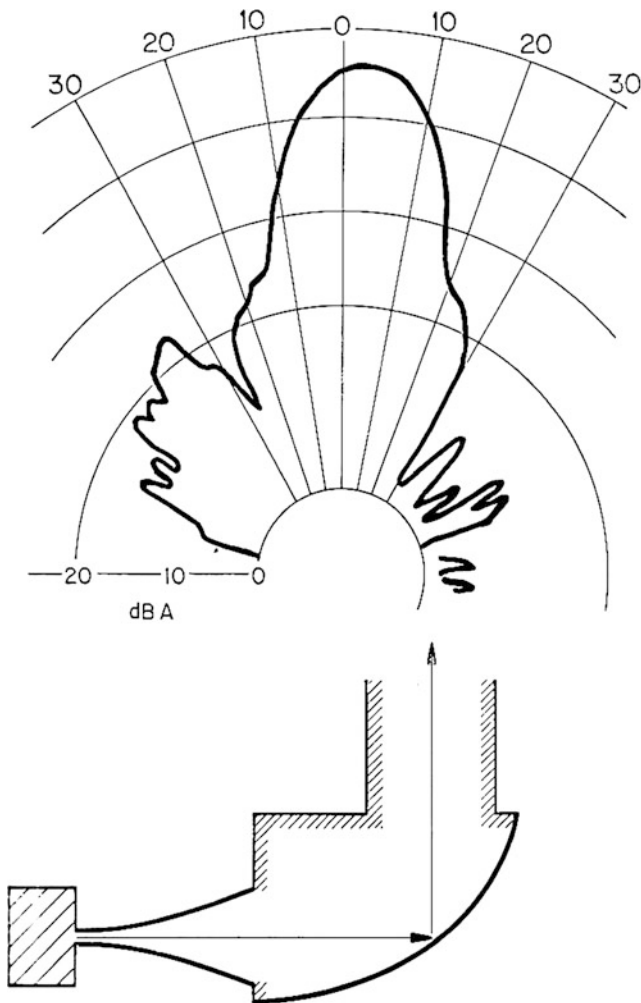
Close to monopoles and dipoles, which are theoretical concepts, is the pulsating sphere of radius a . With the condition that $ka \ll 1$, it is shown (for low displacement velocities) that it is equivalent to the radiation of a dipole source at a sphere's center, with radial velocity at the surface equal to $m \cos(\theta) 2\pi a^3$ with $m = 2\pi a^3 U_0$, where U_0 is the amplitude of the velocity. This analogy between dipoles and pulsing spheres is very useful to solve a large amount of questions relative to acoustic environment such as bubble acoustic emission.

A main question in the domain of acoustic radiation and for remote sensing purposes using transmitters and receivers corresponds to antenna design. Acoustic radiation does not concern very thin beams as in the optical domain but rather larger beams, and it is necessary to know precisely beam angles to determine the observation volumes. Once the acoustic source has been chosen as *microphones*, *loudspeakers*, *transducers*, or *compression chambers*, which convert electricity in pressure fluctuations and acoustic waves, we next choose an antenna to transmit and receive acoustic waves through the considered medium.

Different types of antennas are used (as multi-beam antennas, synthetic antennas combining different elementary sources, horn and parabolic antennas), from which several properties of transmitted and received signals must be reached: *directivity*, which corresponds to being able to get a signal in a preferential direction, *antenna beam control*, which concerns the knowledge and design of antenna beam angle and to limit secondary beams to a very low level, as, for example, for atmospheric sounding with SODAR.

For the first acoustic sounders, parabolic antennas have been mainly used; see Neff and Coulter (1986). Though it is easy to theoretically and analytically model such antennas, it often needs heavy computation; see Rocard (1951). A validation of antenna design in an *anechoic chamber* (a shielded room designed to attenuate wave echoes caused by reflections from the internal surfaces of the room), at least to test far-field behavior, is always necessary to qualify the antenna response. This makes necessary the use of large anechoic chambers to realize far-field conditions. Whatever is the antenna design, a perfect knowledge of antenna beam is necessary and will limit unpleasant surprises as important secondary beams or a beam larger than what was wanted could occur.

In Figure 1, we show experimental results obtained from the design of an *offset antenna* for a 6 kHz acoustic sounder controlled in an anechoic chamber. An offset dish antenna is a type of satellite dish. It is so-called offset because the antenna feed is offset to the side of the



Acoustic Radiation, Figure 2 At the top, antenna beam characteristics as measured in an anechoic chamber. At the bottom, a simple schema of the antenna characteristics showing an exponential horn transmitting sound in a portion of parabola.

reflector, in contrast to a typical circular parabolic antenna where the feed is in front of the center of the reflector.

The antenna's characteristics are shown on Figure 2. Notice that if directivity is good and if secondary beams are relatively low, the antenna beam always differs from theory since it is not easy to model all the elements of the antenna and anyway we have always to validate the modeling. However, a beam width of 13° and a receiver gain of 116–126 dB were obtained, which were in fact the objectives of the design.

Conclusion

We have presented elementary elements about acoustic radiation. Of course, acoustic radiation study in the atmosphere and ocean requires deep specific studies, but considering acoustic remote sensing, questions must be mainly summed up to in regard to radiation pattern of

acoustic sources and to antenna design. However, monopole and dipole behavior and also more complicated concepts such as quadrupoles and higher acoustic poles systems, which are not presented here, suggest what kind of questions have generally to be solved.

Bibliography

<http://mysite.du.edu/~jcalvert/waves/radiate.htm>

- Neff, W. D., and Coulter, R., 1986. Acoustic remote sensing. In Lenschow, D. (ed.), *Probing the Atmospheric Boundary Layer*. Boston, MA: American Meteorological Society, Vol. 201, p. 242.
- Rocard, Y., 1951. *Dynamique générale des vibrations*, 4th edn. Paris: Masson & Cie, 459 pp.
- Vagle, S., and Farmer, D. M., 1992. The measurement of bubble size distributions by acoustical backscatter. *Journal of Atmospheric and Oceanic Technology*, **49**(9), 630–644.
- Weill, A., and Blanc, M., 1987. Design of an acoustic package for Titan boundary layer and surface properties knowledge: the ASTEK system (Acoustic System for Titan Environment Knowledge). In *Proceedings of the 'CASSINI Surface Workshop'*. Washington, DC: E.S.A., 16 pp.

Cross-references

[Acoustic Waves, Propagation](#)
[Acoustic Waves, Scattering](#)

ACOUSTIC TOMOGRAPHY, OCEAN

Brian Dushaw

Applied Physics Laboratory, University of Washington,
 Seattle, WA, USA

Synonyms

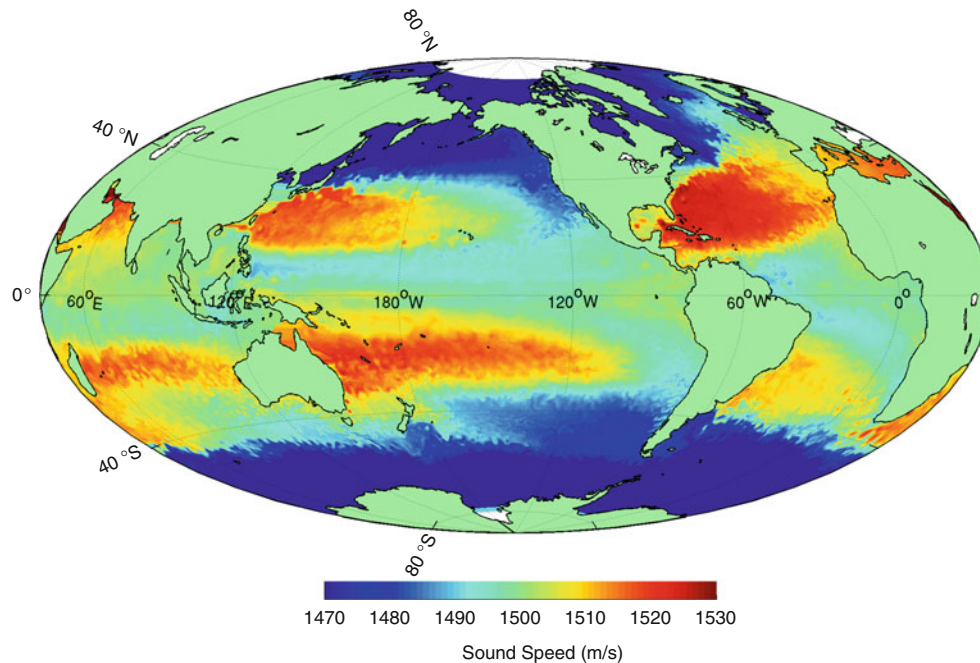
Acoustic thermometry; Acoustic Thermometry of Ocean Climate (ATOC); Moving ship tomography; Ocean acoustic tomography (OAT); Reciprocal tomography

Definition

Ocean acoustic tomography is a remote sensing technique that employs the transmission of sound over large distances within the ocean to precisely estimate averages of temperature and current. Acoustic tomography data usually consist of time-of-flight travel times of acoustic pulses, which represent natural integrating measures of sound speed and current along acoustic paths. Variations in sound speed are predominantly caused by variations in temperature.

Introduction

Acoustic tomography is a technique for measuring large-scale ocean temperature and current using acoustic signals propagating over 100–1,000 km distances. The technique relies on the nature of the oceanic sound speed profile, which acts as an acoustic waveguide, and the transparency of the ocean to low-frequency sound. Sound speed is



Acoustic Tomography, Ocean, Figure 1 Sound speed at 300 m depth derived from an ocean state estimate from the “Estimating the Circulation and Climate of the Ocean” (ECCO) project. Sound speed is a proxy variable for temperature over most of the world’s oceans.

a function of temperature, salinity, and pressure (Jensen et al., 2011), with an approximate value of 1.5 km/s (Figure 1). Over most of the world’s oceans, the sound speed profile has a minimum at about 1,000 m depth, with sound speed that increases towards the surface as a result of increasing temperature and increases towards the seafloor as a result of increasing pressure. Acoustic signals are therefore trapped in the sound channel by refraction (Figure 2). By recording time series of travel times of acoustic signals, the variability of ocean temperature can be inferred. A typical acoustic arrival pattern consists of a set of 5–15 pulse arrivals spanning several seconds, corresponding to a set of distinct acoustic ray paths. Tomography is a unique measurement in that it is inherently averaging: over range along the path of acoustic propagation and over depth from the cycling of the acoustic signals over the water column. The measurements of large-scale ocean temperatures and currents can be extraordinarily precise.

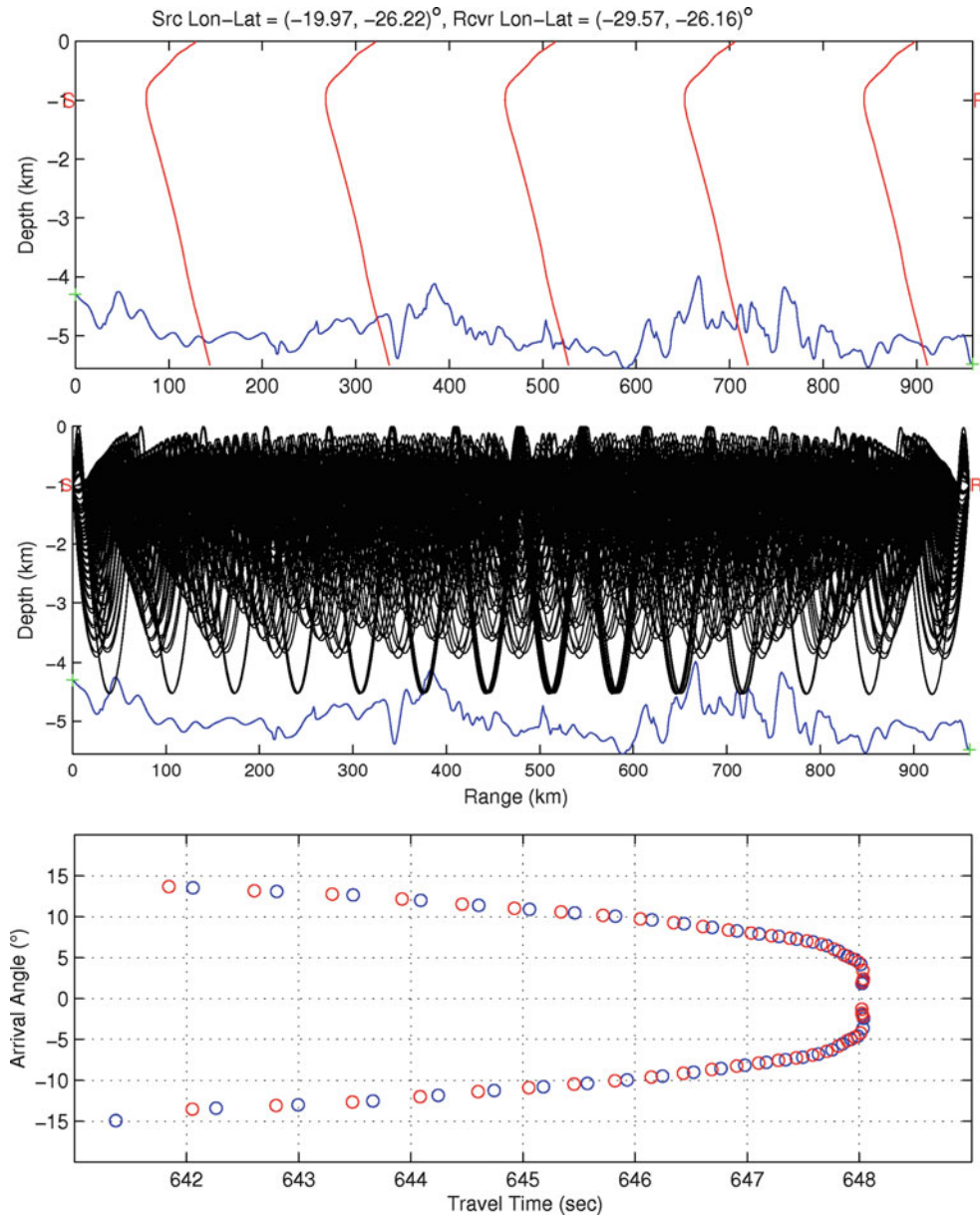
The concept of acoustic tomography was originally proposed by Walter Munk and Carl Wunsch in the late 1970s in response to the discovery that mesoscale variability in the ocean was intense and ubiquitous (Munk and Wunsch, 1982; Munk et al., 1995; von Storch and Hasselmann 2010). The mesoscale presented a challenging observational problem for oceanography. The proposed answer to this challenge was the in situ integrating measurements of acoustic tomography combined with the near-synoptic measurements of sea-surface height by satellite altimetry. Several acoustic sources and

receivers were to be deployed, forming an array with many acoustic paths crisscrossing a region of interest. The term “tomography” was borrowed from medical tomography, to bring to mind those imaging capabilities. Information available from ocean acoustic tomography is much sparser than from medical tomography, however. From the beginning, the acoustic measurements were to be combined with information from other data types using ocean modeling and data assimilation techniques for optimal estimation of the ocean state (Cornuelle and Worcester, 1996; Menemenlis et al., 1997; The ATOC Consortium, 1998; Rémy et al., 2002; Lebedev et al., 2003).

Sound speed and current

Nominally, a 1 °C change in temperature corresponds to a 4 m/s change in sound speed, while a 1 PSU change in salinity corresponds to a 1.3 m/s change in sound speed. Observed changes in sound speed, measured by tomography through the changes in travel times of acoustic pulses, are therefore ambiguous between temperature and salinity. As a practical matter, however, changes in ocean temperature of 1 °C are common, while changes in ocean salinity of 1 PSU are rare. Other than in extreme environments, the possible effects of salinity changes on acoustic travel times can be ignored.

Ocean currents also affect the time of flight of acoustic pulses, although the magnitudes of current variations are



Acoustic Tomography, Ocean, Figure 2 Top: Sound speed profiles derived from the NOAA 2009 World Ocean Atlas along a 960 km section along 26°S in the Brazil Basin. Middle: Associated acoustic ray paths between a source and a receiver, both at 1,000 m depth. Bottom: Associated ray arrival pattern shown as arrival angle versus travel time. The rays arrive in groups of four, corresponding to the possible signs of the ray angles at the source and receiver. The Brazil Basin has almost ideal characteristics for acoustic tomography. Each region of the ocean has its own unique acoustic properties.

usually an order of magnitude smaller than sound speed variations. The two effects can be separated using reciprocal acoustic transmissions. Reciprocal tomography employs transceivers that transmit coincident signals in opposite directions. The travel times of acoustic pulses with the current will be different than the travel times against the current. By forming the sum and difference of reciprocal travel times, the contributions of sound speed and current can be distinguished.

An early test of reciprocal tomography (RTE87) was conducted in 1987 in the central North Pacific using reciprocal acoustic propagation over ranges of O (1,000 km). Current variations of about 10 mm/s were measured, while sound speed variations were about 25 m/s (Munk et al., 1995). Expressed in terms of acoustic travel times, the nominal time of flight of the acoustic pulses in this experiment was about 600 s, the variations in temperature caused travel time variations of about

0.1 s (summertime warming of the near-surface ocean), and the variations in currents caused travel time variations of about 0.005 s (tidal and low-frequency barotropic currents).

As a corollary to the measurement of current, the integration of currents over the paths of a triangular array of tomographic transceivers is a measure of areal-averaged relative vorticity by Stokes' theorem (Munk et al., 1995). The precision of this measurement can be illustrated by the use of tomography to measure "tidal vorticity" or the changes to relative vorticity primarily induced by the changes in water column depth by tidal elevation. Tidal vorticity of order 10^{-9} s^{-1} , five orders of magnitude less than the planetary vorticity (local inertial frequency), was measured in the western North Atlantic using a pentagonal tomographic array of 660 km diameter deployed in 1991 (Dushaw et al., 1997).

Rays and modes

The determination of information about ocean variability from acoustic data is an inverse problem that requires an ocean model that can be fit to the data using weighted least squares techniques (Munk et al., 1995; Worcester, 2001). Possible models range from a simple time-independent model employing a truncated Fourier series and baroclinic modes to represent horizontal and vertical variability to sophisticated time-dependent dynamical models. The choice of the model depends on the nature of the observed ocean variability and the goals of the oceanographic analysis.

The prerequisite for any inverse problem is a solution for the forward problem, however. Predictions for the acoustic arrival pattern of travel times in any particular experiment geometry can be readily computed by a variety of means, including ray tracing, acoustic modes, or the parabolic wave equation (Jensen et al., 2011). Accounting for a variety of acoustic properties, such as ray travel time and arrival angle, identification of predicted arrivals with measured arrivals is usually unambiguous. This identification is often impossible when the acoustic signals have interacted with the sea floor, however. The tomographic information about the ocean is indicated by the small discrepancies between the predictions and measurements. One goal of this identification is to determine the spatial sampling characteristics, or "measurement kernels," of the acoustics. Most often the measurement kernels are taken to be just the ray paths identified with particular ray arrivals. The measured ray travel times are the integrals of the reciprocal of sound speed along the ray paths. The measurement kernel, combined with the ocean model, enables the essential elements of the inverse problem to be computed, and the weighted least squares solution of that inverse problem gives the desired estimate of ocean variability from the acoustic data.

While acoustic rays have been the measurement kernels most commonly employed for the inversion of

tomography data, acoustic modes have also been employed, particularly in polar regions where the polar sound speed profile confines the lowest modes near the surface. The upper ocean is a region of particular interest, of course. The near-surface modes are also sometimes matched to particular water masses in the Arctic (Mikhalevsky and Gavrilov, 2001).

Recently, "travel-time sensitivity kernels" have been computed, which relate the sampling of the complete acoustic wave field to particular travel times (Skarsoulis et al., 2009). Stemming from this rigorous description of the acoustic field was the important proof, long known in practice, that the measurement kernels computed using the geometric-ray approximation are an accurate representation of the actual sampling associated with particular ray arrivals.

Applications

Over the past 30 years, tomography has been employed for wide-ranging applications (Munk et al., 1995; Dushaw et al., 2010). Several examples that highlight the strengths and roles of the measurement will be mentioned here; more substantive reviews of applications can be found in Munk et al. (1995), Dushaw et al. (2001, 2010), and Worcester (2001).

The Greenland Sea Project deployed a six-mooring tomographic array with 100 km diameter in 1988 in the region of deep convective mixing in the Greenland Sea (Morawitz et al., 1996). The Greenland Sea is one of the few regions where deep water of the world's oceans is formed. The aim of the Greenland Sea Project was to quantify this water formation. Tomography was ideal for this measurement because its remote sensing capability was essential in this harsh, sometimes ice-covered, environment, the quantity measured was the net deepwater formation which is an integrated quantity, and the deepwater formation is episodic and unpredictable in time and location. The rapid sampling of integrated temperature afforded by the acoustic measurements proved to be essential in estimating the net deepwater formation over the winter of 1988/1989. Concurrent, extensive measurements by CTD casts proved to be inadequate to the task.

With the ability to make repeated, integrated measurements along an acoustic path, tomography is often employed to monitor temperature or current averaged across straits or other constricted regions. Mass transport and heat content through the Strait of Gibraltar were measured by reciprocal tomography in 1996 (Send et al., 2002). Temperature variations in Fram Strait are presently monitored by tomography by the multiinstitutional ACOBAR collaboration. Within Fram Strait the complicated West Spitsbergen and East Greenland Current systems transport heat and salt between the North Atlantic and Arctic Oceans. One experiment conducted in the late 1990s by the French Research Institute for Exploration of the Sea (IFREMER) aimed to measure the net transport of heat and salt by subsurface

salt lenses, or Meddies, out of the Mediterranean into the North Atlantic.

With its inherent averaging properties, tomography can make unique, accurate measurements of large-scale barotropic currents. The barotropic currents estimated from the RTE87 data had an uncertainty of about 1 mm/s. Relative vorticity associated with the RTE87 currents was of order 10^{-8} s^{-1} (Munk et al., 1995). Measurements of barotropic tidal currents by tomography are the most accurate available; they have been used to test global tidal models. These capabilities have been underutilized in ocean observation.

Whereas the difference of reciprocal travel times is primarily a measure of depth-averaged current, the sum of reciprocal travel times is primarily a measure of the sound speed signature of the first internal or baroclinic mode. The exact resolution of vertical variability, or resolution of higher-order modes, depends on the available vertical sampling of the rays in any particular region. The inherent depth average of the measurement makes the first baroclinic mode the dominant signal, however, when ocean variability is characterized by baroclinic modes. This property led to the detection of radiation of coherent mode-1 internal tides far into the ocean's interior and placed tomography at the forefront of internal-tide and ocean mixing revolution that has unfolded over the past two decades. One interesting aspect of these observations is that an acoustic path acts as line-segment antenna for the internal-tide radiation. The beam pattern for this antenna has narrow maximum response for wave numbers perpendicular to the acoustic path, corresponding to wave crests aligned along the acoustic path. The culmination of these observations, brought together with the more synoptic observations of these waves by satellite altimetry, was the demonstration that mode-1 internal tides propagate coherently across ocean basins. Tidal harmonic constants can be used to predict the amplitude and phase of the mode-1 internal tides over many regions of the world's oceans (Dushaw et al., 2011).

Over the past 30 years, there have been numerous experiments employing multipath tomography arrays for ocean observation. These experiments have taken advantage of the fact that the number of paths of an array increases quadratically with the number of deployed instruments. Experiments have often consisted of a 5 or 6 mooring pentagonal array, as was the case in the Greenland Sea array (Morawitz et al., 1996), the 1990–1991 Acoustic Mid-Ocean Dynamics Experiment (AMODE) in the western North Atlantic (Cornuelle and Worcester, 1996), and the 1997 Kuroshio Extension Pilot Study array (Lebedev et al., 2003). Other experiments have been deployed in the equatorial Pacific (Kaneko et al., 1996). The 2000–2001 Central Equatorial Pacific Tomography Experiment (CEPTE) was aimed at observing the weak meridional currents of the equatorial current system. These experiments have been designed following the original Munk and Wunsch notion to better understand a complicated region by making sparse tomographic

observations that are combined with all available ancillary data by ocean modeling and data assimilation techniques. These studies have quantitatively shown that the tomographic data type affords a significant resolution of ocean properties not possible by other data types.

Acoustic thermometry of ocean climate

The use of acoustic transmissions across ocean basins to measure the temperature has come to be called acoustic thermometry. These measurements are made possible by lowering the transmitted acoustic frequency to reduce the sound attenuation. With 20–60 Hz sound signals, the range of acoustic propagation in the ocean does not appear to be limited. The aim of these basin-scale observations is to precisely quantify the large-scale changes in ocean temperature (Dushaw et al., 2001, 2010).

One of the first tests of long-range acoustic transmissions for the purpose of ocean climate measurements was the Heard Island Feasibility Test (HIFT). In a 9 day test in 1991, 57 Hz acoustic signals were transmitted from an array of acoustic sources lowered from the R/V Cory Chouest near Heard Island in the southern Indian Ocean (Munk and Baggeroer, 1994). The site is a location with unblocked acoustic paths to both coasts of the United States, and, indeed, the transmitted signals were recorded off the coasts of Nova Scotia and Washington state. The attempts to measure ocean temperatures over antipodal acoustic ranges were then abandoned with the recognition that measurements across several climatological regimes were perhaps less useful than basin-scale measurements.

The HIFT was succeeded by the decade-long (1995–2006) Acoustic Thermometry of Ocean Climate (ATOC) program (The ATOC Consortium, 1998; Dushaw et al., 2009). This program deployed 75 Hz acoustic sources off central California and north of Hawaii to transmit tomographic signals across the North Pacific Basin. The program employed receivers of opportunity and arrays of hydrophones specifically designed to determine the properties of acoustic propagation over several megameter ranges. The acoustic data were used to test accuracy of ocean state estimates of the North Pacific obtained by various means, including simple forward integration of a model, objective analysis of hydrographic and altimeter data, and assimilation of available data to constrain general circulation models. The travel times measured over a decade were compared with equivalent travel times derived from the several state estimates. The comparisons of computed and measured time series provided a stringent test of the accuracy of the large-scale temperature variability in the models. The differences were sometimes substantial, indicating that acoustic thermometry data does provide significant additional constraints for numerical ocean models.

Averages of temperature across the Arctic Ocean were measured in the 1994 Transarctic Acoustic Propagation (TAP) and 1999 Arctic Climate Observations Using Underwater Sound (ACOUS) experiments (Mikhalevsky

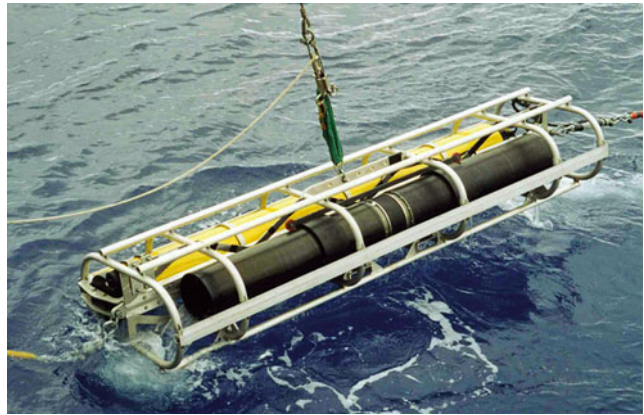
and Gavrilov, 2001). These experiments were among the first observations of warming of the Atlantic Layer in the Arctic Ocean. Using acoustics to remotely sense ocean temperature and other properties under the ice is a particularly compelling application of tomography.

Instrumentation

One of the technical advances that made tomography possible was the development of broadband, controlled sound sources in the 1970s to replace explosive charges. By employing lengthy coded signals and pulse compression techniques, the energy of the sound signals is spread over several minutes, while the travel time resolution achieved after signal processing is about a millisecond. The peak pressure of these controlled signals is much less than signals from the explosive sources. Typical sound levels are 25–250 W (or 185–195 dB re 1 μ Pa at 1 m). Low-frequency (75–300 Hz), broadband acoustic sources are large and heavy, however. The size of the source is determined by the acoustic wavelength, which, for 250 Hz signals, is 6 m. One acoustic source commonly used today is a tuneable organ-pipe transducer of about 4 m length and 1,000 kg weight that usually broadcasts a swept-frequency signal from 200 to 300 Hz (Figure 3). Broadband acoustic sources, such as the HLF-5, are also power hungry. Moored acoustic sources powered by lithium or alkaline battery packs can transmit coded signals of a few minutes length several times a day for durations of about a year. Timekeeping is maintained using a rubidium atomic clock. The 75 Hz acoustic sources of the ATOC program were mounted on the sea floor and powered and controlled by a cable to shore.

A tomographic receiver typically consists of a small vertical array of about 4 hydrophones spaced by 1.5 wavelengths. Hydrophone arrays allow for beamforming of the received signals, which boosts the SNR of the acoustic arrival pattern and gives a determination of the ray arrival angles. The arrival angle information is important for matching recorded to predicted arrivals and for distinguishing rays that arrive at about the same time but with different angles (Figure 2, bottom). Tomography transceivers usually consist of a mooring with an acoustic source placed near the sound channel axis and a small vertical line array (VLA) immediately above or below. Both instruments are controlled by a single electronics package.

Recently hydrophone modules were developed with the ability to communicate with a controlling electronics package by induction along the mooring wire. These hydrophone modules allow the deployment of VLAs of any number and spacing, since they can be clamped on the mooring wire at arbitrary locations. For acoustics research, VLAs of 100 or more hydrophones have been employed. These modules include thermistors as well, so that a VLA can act as a thermistor chain. Hydrophones have a broadband sensitivity and of course can also be used for studies of ambient sound.



Acoustic Tomography, Ocean, Figure 3 A Webb Research swept-frequency acoustic source as it is deployed from a ship. The black cylinder is the “organ pipe” acoustic source, and the yellow cylinder is an alkaline battery pack. In the photo the source is horizontal; once deployed the right hand side will be oriented towards the surface. The source is mounted within an aluminum cage for protection during shipping and deployment. Courtesy of Lloyd Green, Scripps Institution of Oceanography.

One extension of tomography is the concept of “Moving Ship Tomography.” The technique aims to make the most of a deployed array of acoustic sources by circumnavigating this array while repeatedly lowering a hydrophone array from a ship to receive the acoustic signals. The aim is to eventually accumulate enough acoustic data that the temperature field within the circumnavigation can be mapped to considerable resolution. Since it can take up to 60 days to circumnavigate a region of 1,000 km diameter, data assimilation techniques are required to account for the evolution of the temperature field while the data are obtained. Recently hydrophones have been experimentally deployed on gliders to receive signals transmitted by tomographic sources.

Passive tomography is an experimental technique that aims to avoid the expense and trouble of moored acoustic sources by using ambient sounds as tomographic signals. The technique employs coordinated receiving arrays deployed on either side of an area of observation. Ambient sound, preferably from such sources as a distant ship or whale, propagating past one array then forms a known signal when it is received on the second array. By comparing the two acoustic signals, information about the intervening ocean might be inferred.

Marine mammals and active acoustics

The HIFT and ATOC programs engendered considerable controversy concerning the possible effects of the transmitted sounds on marine mammals and other marine life (Potter, 1994). The issue presented a formidable challenge in public relations, since it included marine science, climate, and acoustics topics, many of which are still areas of active research. For example, one newspaper report

confused sound levels in air with those in water, which are defined differently, making the underwater acoustic sources appear to be several orders of magnitude louder than they actually were. The sound transmitted by the ATOC sources was 195 dB re 1 μ Pa at 1 m (about 250 W), with a low duty cycle (ca. 2 %) consisting of eight brief transmissions every few days. Because of spherical spreading, 1,000 m from the source the signal level was only 135 dB (about 2.5×10^{-4} W).

As a result of the controversy, the 1996–2006 ATOC and NPAL (North Pacific Acoustic Laboratory) projects included a Marine Mammal Research Program (MMRP) to study the potential effects, if any, of the ATOC sound sources on marine mammals and other marine life. (Indeed, the issue of the effects of acoustics on marine mammals proved to be a boon for marine mammal research.) The MMRP did not find any overt or obvious short-term changes in the distribution, abundance, behavior, or vocalizations of marine mammals in response to the playback of ATOC-like sounds or in response to the transmissions of the ATOC sound sources themselves. The MMRP investigators concluded that the ATOC acoustic transmissions caused no significant biological impact (National Research Council, 2000).

Summary

Ocean acoustic tomography uses the travel times of coded acoustic signals transmitted over 100–1,000 km ranges to infer precise information about the ocean temperatures and currents. In most regions of the world's oceans, sound speed is an accurate proxy variable to temperature; salinity contributions are negligible. One primary strength of the measurement is that it is inherently averaging, suppressing the small-scale internal wave or mesoscale noise that can make it difficult to observe the large-scale signal. Predicted acoustic ray paths can be identified with measured multipath arrivals and used in an inverse analysis to estimate ocean variability. Ray paths have been shown to accurately represent the measurement kernel for the recorded arrivals. The variations of both sound speed and current can be estimated using acoustic transmissions sent reciprocally between transceiver pairs.

Tomography data are perhaps best employed by combining them with other data types, such as satellite altimetry and Argo floats, using data assimilation techniques and ocean models. Such techniques bring together dynamical constraints and disparate data types, each with its own sampling, bias, and error characteristics, to obtain optimized state estimates for the ocean.

Over the past 30 years, tomography has been employed for disparate applications, ranging from the quantification of deepwater formation in the Greenland Sea to the detection of internal-tide radiation from the Hawaiian Ridge to basin-wide measurements of temperature in the central North Pacific. Tomography often requires careful analysis and interpretation, depending on the unique acoustic and oceanographic conditions of the region where it is

employed. Whether used for regional- or basin-scale observations, acoustic remote sensing has been quantitatively shown to provide information about ocean variability that is not possible to obtain by other approaches.

Acoustic tomography was accepted as part of the emerging Ocean Observing System during both of the OceanObs'99 and '09 international workshops (Dushaw et al., 2001; Dushaw et al., 2010). Within the observing system context, the instrumentation for tomography can serve multiple purposes. Hydrophone arrays are used to study a wide range of human, biological, and geological activity. Acoustic sources can transmit signals that can serve other purposes, such as signals that can be used to track drifting instrumentation. Thus, a modest set of active and passive acoustic instrumentation deployed worldwide can form a general-purpose global acoustic observing network (Howe and Miller, 2004; Boyd et al., 2011).

Bibliography

- Boyd, I. L., Frisk, G., et al., 2011. An international quiet ocean experiment. *Oceanography*, **24**, 174–181, doi:10.5670/oceanog.2011.37.
- Cornuelle, B. D., and Worcester, P. F., 1996. Ocean acoustic tomography: integral data and ocean models. In Malanotte-Rizzoli, P. (ed.), *Modern Approaches to Data Assimilation in Ocean Modeling*. New York: Elsevier, pp. 97–115.
- Dushaw, B. D., Egbert, G. D., Worcester, P. F., Cornuelle, B. D., Howe, B. M., and Metzger, K., 1997. A TOPEX/POSEIDON global tidal model (TPXO.2) and barotropic tidal currents determined from long-range acoustic transmissions. *Progress in Oceanography*, **40**, 337–367.
- Dushaw, B. D., et al., 2001. Observing the ocean in the 2000s: a strategy for the role of acoustic tomography in ocean climate observation. In Koblinsky, C. J., and Smith, N. R. (eds.), *Observing the Oceans in the 21st Century*. Melbourne: GODAE Project Office and Bureau of Meteorology, pp. 391–418.
- Dushaw, B. D., Worcester, P. F., et al., 2009. A decade of acoustic thermometry in the North Pacific Ocean. *Journal of Geophysical Research*, **114**, C07021, doi:10.1029/2008JC005124.
- Dushaw, B. D., et al., 2010. A global ocean acoustic observing network. In Hall, J., Harrison, D. E., and Stammer, D. (eds.), *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*. Venice: ESA Publication WPP-306, Vol. 2.
- Dushaw, B. D., Worcester, P. F., and Dzieciuch, M. A., 2011. On the predictability of mode-1 internal tides. *Deep Sea Research*, **58**, 677–698, doi:10.1016/j.dsr.2011.04.002.
- Howe, B. M., and Miller, J. H., 2004. Acoustic sensing for ocean research. *Marine Technology Society Journal*, **38**, 144–154.
- Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H., 2011. *Computational Ocean Acoustics*, Second edn. New York: Springer, doi:10.1007/978-1-4419-8678-8. ISBN 978-1-4419-8677-1.
- Kaneko, A., Zheng, H., Nakano, I., Yuan, G., Fujimori, H., and Yoneyama, K., 1996. Long-term acoustic measurement of temperature variations in the Pacific North Equatorial Current. *Journal of Geophysical Research*, **101**, 16,373–16,380, doi:10.1029/96JC01053.
- Lebedev, K. V., Yaremchuk, M., Mitsudera, H., Nakano, I., and Yuan, G., 2003. Monitoring the Kuroshio Extension through dynamically constrained synthesis of the acoustic tomography, satellite altimeter and in situ data. *Journal of Oceanography*, **59**, 751–763.

- Menemenlis, D., Webb, A. T., Wunsch, C., Send, U., and Hill, C., 1997. Basin-scale ocean circulation from combined altimetric, tomographic and model data. *Nature*, **385**, 618–688.
- Mikhalevsky, P. N., and Gavrilov, A. N., 2001. Acoustic thermometry in the Arctic Ocean. *Polar Research*, **20**, 185–192, doi:10.1111/j.1751-8369.2001.tb00055.x.
- Morawitz, W. M. L., Sutton, P. J., Cornuelle, B. D., and Worcester, P. F., 1996. Three-dimensional observations of a deep convective chimney in the Greenland Sea during winter 1988/1989. *Journal of Physical Oceanography*, **26**, 2316–2343.
- Munk, W., and Baggeer, A., 1994. The Heard Island papers: a contribution to global acoustics. *Journal of the Acoustical Society of America*, **96**, 2327–2329, doi:10.1121/1.411316.
- Munk, W., and Wunsch, C., 1982. Observing the ocean in the 1990s. *Philosophical Transactions of the Royal Society London A*, **307**, 439–464, doi:10.1098/rsta.1982.0120.
- Munk, W., Worcester, P., and Wunsch, C., 1995. *Ocean Acoustic Tomography*. Cambridge: Cambridge University Press. ISBN 0-521-47095-1.
- National Research Council, 2000. *Marine Mammals and Low-Frequency Sound: Progress Since 1994*. Washington, DC: National Academy Press. 160 pp, http://www.nap.edu/catalog.php?record_id=9756. ISBN ISBN-10: 0-309-06886-X.
- Potter, J. R., 1994. ATOC: sound policy or enviro-vandalism? Aspects of a modern media-fueled policy issue. *The Journal of Environment & Development*, **3**, 47–62, doi:10.1177/107049659400300205.
- Rémy, E., Gaillard, F., and Verron, J., 2002. Variational assimilation of ocean tomographic data: twin experiments in a quasi-geostrophic model. *The Quarterly Journal of the Royal Meteorological Society*, **128**, 1739–1458, doi:10.1002/qj.200212858317.
- Send, U., Worcester, P., Cornuelle, B. D., Tiemann, C. O., and Baschek, B., 2002. Integral measurements of mass transport and heat content in the strait of Gibraltar from acoustic transmission. *Deep Sea Research Part II*, **49**, 4069–4095, doi:10.1016/S0967-0645(02)00143-1.
- Skarsoulis, E. K., Cornuelle, B. D., and Dzieciuch, M. A., 2009. Travel-time sensitivity kernels in long-range propagation. *The Journal of the Acoustical Society of America*, **126**, 2223–2233, doi:10.1121/1.3224835.
- The ATOC Consortium, 1998. Ocean climate change: comparison of acoustic tomography, satellite altimetry, and modeling. *Science*, **28**, 1327–1332, doi:10.1126/science.281.5381.1327.
- von Storch, H., and Hasselmann, K., 2010. *Seventy Years of Exploration in Oceanography: A Prolonged Weekend Discussion with Walter Munk*. Berlin: Springer-Verlag, doi:10.1007/978-3-642-12087-9. ISBN 978-3-642-12086-2.
- Worcester, P. F., 2001. Tomography. In Steele, J., Thorpe, S., and Turekian, K. (eds.), *Encyclopedia of Ocean Sciences*. San Diego: Academic Press, pp. 2969–2986.

Links

- Acoustic Technology for Observing the interior of the Arctic Ocean (ACOBAR): <http://acobar.nersc.no>
- The Heard Island Feasibility Test (HIFT): <http://909ers.apl.washington.edu/dushaw/heard/>
- North Pacific Acoustic Laboratory (NPAL): <http://npal.ucsd.edu/>
- Ocean Acoustics Library: <http://oalib.hlsresearch.com/>
- Discovery of Sound in the Sea (DOSITS) <http://www.dosits.org/>
- Ocean acoustic tomography on Wikipedia: http://en.wikipedia.org/wiki/Ocean_acoustic_tomography
- A day in the life of a tomography mooring: <http://staff.washington.edu/dushaw/mooring/>

ACOUSTIC WAVES, PROPAGATION

Alain Weill

Bur. Jussieu, LATMOS, Laboratoire Atmosphere Milieux Observations Spatiales, Paris, France

Synonyms

Acoustic propagation; Sound waves propagation

Definition

Propagation. Concerns acoustical wave transmission inside a solid, liquid, or gaseous physical medium without material entrainment.

Introduction

After a very simple description of acoustical geometric propagation and plane wave assumption at large distance from sources, we shall speak about acoustic absorption by the medium as a function of frequencies and medium parameters. Then we shall present Doppler shift, which is very fundamental for remote sensing techniques. We shall be in a position to discuss sound propagation in a stratified media having some mean properties variation as a function of distance of acoustic sources or receivers. This last question is very important for atmosphere and ocean remote sensing where acoustic ray tracing has to be performed to know how acoustic waves propagate and from what kind of target an acoustic signal comes.

Wave propagation

If a pure sound corresponds to a sinusoidal wave, a sound can generally be considered as a superposition of sinusoidal waves of different frequencies f_i .

At the acoustic source 0, we can say that the sound amplitude is

$$y(0, t) = \sum_i (a_i \sin(2\pi f_i t)) \quad (1)$$

If the different frequencies at the origin are considered to be in phase, c the wave velocity is supposed not to depend on frequency.

In the entry on “[Acoustic Radiation](#),” we have described elementary acoustic sources as monopoles, dipoles, and pulsating spheres. Let us consider an acoustical source at one point, which radiates spherical waves. The acoustic intensity at different distances remains the same, considering all of the sphere points at distance r from the source (repartition in the surface $4\pi r^2$), but the acoustic intensity at each point, at the distance r will be divided by $4\pi r^2$. This corresponds to what is called *spherical divergence*. As an easy rule, if the distance is multiplied by 10, the sound intensity will be 100 times smaller.

Notice that far from the source, the wave can be considered as a plane wave.

Wave absorption

Let us consider, as an example, an air layer of depth d and a sound of frequency f_i propagating across this layer. Several observations can be noticed:

If the incident sound has an incident intensity I_{inc} , one part of the sound I_{ref} can be reflected by the layer, one part I_{abs} can be absorbed, and finally one part is transmitted I_{tr} . For the depth d , we can say that the air absorption coefficient corresponds to $a = I_{abs}/I_{inc}$. This absorption coefficient corresponds to absorption at the acoustic frequency f_i .

In the air, sound absorption depends on frequency, temperature, and humidity. A very important reference is Harris (1966), who has performed several laboratory experiments to measure sound attenuation in air (in dB/100 m) as a function of frequency, which is very useful to choose power and frequencies to design acoustic sounders. As the behavior as a function of frequency, of temperature, and humidity is not linear, it is preferable here just to indicate behavior elements and to refer to Harris' (1966) work for precise attenuation evaluation.

It has to be remarked that attenuation grows with frequencies and it presents a maximum, which increases with humidity in percent (maximum of attenuation close to 1 dB/100 m at 2,000 Hz for a typical humidity of 5 % and a maximum of 8.5 dB/100 m for a 20 % of humidity at 25,000 Hz for a temperature of 20 °C).

For humidity between 30 % and 90 %, frequency variation is rather linear (0.3 dB/100 m for 2,000 Hz and 0.8 dB/100 m for 4,000 Hz). Particularly important is to consider that for a chosen frequency, attenuation as function of temperature and humidity in percent presents a maximum, which increases when humidity decreases. Typically, for low humidity below 30 %, attenuation increases as a function of humidity and temperature and for larger temperature (typically higher than 30°), the behavior is reverse and attenuation decreases for higher humidity.

For the ocean, sound absorption is smaller than in the air. It depends on salinity, temperature, and pH (acidity) of the ocean.

Here r is the distance and I_0 is acoustic power transmitted before attenuation. It is close to 0.011 dB/100 m at 2,000 Hz at a 100 m depth. Notice that if the same frequencies as in air are considered for the ocean, wave lengths are, of course, five times larger due to sound speed.

Doppler shift

This very general phenomenon, which is transversal to many remote sensing domains, (electromagnetic, optical, and acoustic), was discovered at the end of the nineteenth century.

For acoustic waves, it corresponds to a frequency shift between a transmitter and a receiver when distances between both change as a function of time. This effect was discovered by an Austrian physicist Christian Doppler in 1842 and confirmed from experiments performed by the Dutch meteorologist Buys Ballot in 1845. Buys Ballot has verified Doppler shift with a special

experimental protocol. Musicians were playing horns on a train of known speed and trained musicians on the railway station were hearing sound changes due to train speed.

If f was the transmitted horn frequency played by musicians, f' the frequency observed by fixed musicians in the railway station, c the sound speed, then the Doppler frequency shift due to the train speed can be written as

$$\Delta f/f = (f - f')/f = v/c$$

where v is train speed.

The observed frequency $f' = f + (fv/c)$ is of course dependent on the relative train velocity.

The Doppler shift is very useful for Doppler Sodar or Sonar to be able, using *distance gates* (surface acoustically illuminated at different distances), to analyze medium properties relatively to observed gates. In fact, with these kinds of instruments, mainly the Doppler spectrum is computed to analyze the mean Doppler shift inside different gates. Radial velocities of the medium, in the case where transmitter and receiver are at the same location (for monostatic systems), are measured.

Reflection and refraction of acoustic waves

It corresponds to a general law of propagation of waves corresponding to the Snell-Descartes- Huygens law used for optical waves. However, ocean and atmosphere are *stratified* in temperature, salinity, current for ocean and in temperature, wind, humidity for atmosphere, which corresponds to the fact that we cannot strictly speak of layers since there are more or less continuous variations of these parameters. Let us consider sound reflection in the ocean. In the ocean, the sound velocity mainly depends on the depth z and one supposes that the surface and the bottom of the ocean are (for a first approximation) horizontal planes. One uses generally the concept of ray approximation, and questions which have to be solved correspond to ray tracing. One necessary (but not sufficient condition) to use this concept is that the relative gradient of sound velocity and the wave length must satisfy (see Brekhovskikh and Lysanov, 1990) the condition:

$$\lambda/c(\partial c/\partial z) \ll 1 \quad (2)$$

We, therefore, apply Snell's law $\cos(\theta)/c(z) = \text{constant}$, where θ is the grazing angle (incidence angle) to vary in thin layers 1,2, ... n , $n+1$, where θ_n , θ_{n+1} , c_n , c_{n+1} are, respectively, grazing angles and sound speeds for layers n , $n+1$.

$$\cos(\theta_{n+1})/c_{n+1} = \cos(\theta_n)/c_n \quad (3)$$

and

$$\cos(\theta_{n-1})/c_{n-1} = \cos(\theta_n)/c_n \quad (4)$$

and so on.

When n tends to infinity, then Snell's law is obtained.

With this type of *discretization* (transformation into small layers) of the atmosphere and ocean, one can obtain

propagation in terms of rays and grazing angles tracing. Notice that layers can be obtained considering variations of the main different ocean parameters (salinity, temperature, current) and for the atmosphere (temperature, wind, humidity).

For a considered layer, we get:

$$\theta = [2(c_2 - c_1)/c_1]^{0.5} \quad (5)$$

Therefore, if c_2 is close to c_1 , θ is very small.

To determine the relationship between the ray curvature and sound velocity gradient, we can use Snell's law in the form

$c_0 \cos(\theta) = c \cos(\theta_0)$, where c_0 and θ_0 correspond to values at a fixed (height or depth).

Differentiating this equation with respect to z gives

$d\theta/ds = -\cos(\theta_0)/c_0 \, dc/dz$ with ds the ray element $= dz/\sin(\theta)$

$\cos(\theta_0)/c_0$ corresponds to the ray path element at a known depth or height.

$\text{Abs}[ds/d(\theta)]$ is the *curvature* ray R .

To find rays across the atmosphere and the ocean is related to different propagation computer programs, which have been mainly developed for ocean acoustics.

It is important to say a word about sound channel, the SOFAR channel discovered by Ewing and Worzel (1948). At a depth which varies with geographical location at about 1 km of depth, rays tend to bend toward regions of smaller sound speed and are channeled. Refracted rays can extend toward thousands of kilometers in range without touching surface or bottom. This propagation channel works like the ionosphere for electromagnetic propagation and plays a fundamental role for low-frequency acoustic communication; see Munk (2002). It has to be noticed that ocean and/or ocean currents modifications, which can be related to climatic variations, can have an impact on this propagation channel. This justifies acoustic studies of global warming, see Munk and Forbes (1989).

For the atmosphere, stratification related to inversion layers, fronts or stratosphere can play a similar propagation layer for acoustic waves as for ocean, but temporal and spatial variability is larger and we cannot speak of a "permanent channel."

Conclusion

Propagation of acoustic waves in ocean and atmosphere knowledge is mainly dependent on properties of these mediums. Acoustic remote sensing is indeed used to analyze atmosphere and ocean behavior or acoustic targets behavior inside these domains as a Doppler shift related to currents or wind, and/or parameters which influence acoustic propagation. Particularly, acoustic *tomography* (analysis by layers (*tomos* in Greek), which is used to invert ocean or atmosphere properties) needs a perfect knowledge of ray tracing modeling with basic principles of ray tracing similar to those shown in this entry.

Bibliography

- Brekhovskikh, L. M., and Lysanov, Y. P., 1990. *Fundamentals of Ocean Acoustics*, 2nd edn. Berlin: Springer, 270 pp.
- Ewing, M., and Worzel, J. L., 1948. *Propagation of Sound in the Ocean*. New York: Geological Society of America, Memoir 27.
- Flatté, S. M., Dashen, R., Munk, W. M., Watson, K. M., and Zachariasen, F., 1979. *Sound Transmission Through a Fluctuating Ocean*. Cambridge: Cambridge University Press, 299 pp.
- Harris, C. M., 1966. Absorption of sound in air versus humidity and temperature. *Journal of Acoustical Society of America*, **40**(1), 148–159.
- Munk, W., 2002. Acoustic tomography. In *Encyclopedia of Global Environmental Change*. Chichester: Wiley Editor, Vol. 1, p. 161.
- Munk, W. H., and Forbes, A. M. G., 1989. Global ocean warming: an acoustic measure? *Journal of Physical Oceanography*, **19**, 1765–1778.

Cross-references

[Acoustic Tomography, Ocean](#)

[Electromagnetic Theory and Wave Propagation](#)

ACOUSTIC WAVES, SCATTERING

Alain Weill

Bur. Jussieu, LATMOS, Laboratoire Atmosphere Milieux Observations Spatiales, Paris, France

Definition

Scattering. A way by which obstacles or medium fluctuations of small dimensions can modify acoustic wave propagation in the medium. The difference between scattering processes and the reflection/refraction process is related to acoustic wavelengths. For scattering, acoustic wavelengths are of the order of obstacles and medium fluctuation dimensions. Francisco Maria Grimaldi first described scattering of optical waves in 1665, in the opus *Physicomathesis of lumine*, but this principle was renounced by Fresnel at the end of seventeenth century.

It is important to differentiate between scattering and *diffraction*. Diffraction concerns perturbations whose dimensions are large compared to a wavelength.

Introduction

In the first section, we will describe acoustic scattering by simple obstacles such as spheres. The second section will introduce Lambert scattering and Marsh scattering from surfaces; this process is particularly important for sonar sounding. The third section will introduce acoustic scattering in a turbulent medium; which is useful for understanding propagation in the turbulent atmosphere and ocean. This last process is very fundamental to understanding Sodar (sound detection and ranging) acoustic sounding.

Acoustic scattering of a plane wave on a rigid sphere

Suppose that a plane wave (progressive wave) intercepts a fixed and rigid (not compressible) sphere. As the obstacle is rigid, there is no acoustic emission by the sphere itself and the flow is modified just around the sphere. Here, we use a simple concept of acoustic equivalence to describe acoustic perturbation.

This concept can be summarized as follows.

We examine a small volume (of any shape), and in this case, it is a sphere of shape dimension much less than a wavelength in size. When an acoustic wave intersects the volume, the volume does not transmit any sound since it is not deformed, but it has effects on sound radiation, and sound energy from the incident radiation is scattered in all directions. As in optics, the scattered amplitude is proportional to the wave number at power 4.

1. We have a first perturbation equivalent to a simple acoustic source (monopole) at the center of the sphere that would source and sink sufficient volume to cancel the changes that would occur in its absence.
2. A second perturbation is the same as for a sphere (dipole) oscillating in a medium at rest with a velocity opposite to that of the wave.

Therefore, we can replace the sphere by a single (monopole) and a double source (dipole) located at its center, which has the same effect on the flow at large distances as the sphere did. We assume that $2\pi a/\lambda$ is $\ll 1$, which is equivalent to say that plane wave properties remain constant across the sphere.

This last strict condition on the wavelength and sphere radius has great importance in the choice of wavelengths and frequencies or perturbation dimensions.

The equivalent monopole and dipole will radiate energy in all directions; therefore, the energy in the plane wave will be partially scattered, and we have to evaluate the strength of the equivalent sources.

The change in the incident source is the *condensation* s , which corresponds to the change in volume per unit volume, so the volume compressed into the volume of the sphere is Qs , where Q is the volume of the sphere $4\pi a^3/3$. This would be roughly the same equivalent change for a body of any shape. Hence, the strength of the monopole is the time derivative of this, which gives a strength of $q = -k^2 Q \varphi_0$, q is the strength of the monopole, k the wave number and φ_0 is the source intensity, since the condensation is the derivative of the velocity potential, divided by square of sound velocity in the motion equation. Note that the difference between spatial and temporal derivative is in the factor jk , which makes the difference between temporal and spatial derivative from the wave equation.

The resultant velocity potential from the two equivalent sources acting simultaneously is

$$\begin{aligned} \varphi = & -(ka)^2(2\pi a\varphi_0/3)(2 + 3 \cos(\theta)) \\ & \times (\exp(-jkr)/r) + (ka)(2\pi a\varphi_0) \cos(\theta) \\ & \times \exp(-jkr)/r^2 \end{aligned} \quad (1)$$

where θ is the angle between r and the direction of the acoustic source. The first term is dominant for large distances, since the second term, corresponding to near field, vanishes rapidly.

Consider now the scattered intensity derived from the acoustic pressure; it is proportional to:

1. The intensity of the incoming source
2. $(ka)^4$
3. The sphere surface cross-section
4. $4 + 12\cos(\theta) + 9 \cos(\theta)^2$

This is similar as for electromagnetic waves scattering, except for angular dependence.

Integration in all θ shows an equivalent result, as if the dipole and the monopole were independently radiating. (π in radians is equal to 180°)

If A is the amplitude of the displacement, we get a total scattered intensity:

$$W = (\pi a^2)(\rho c \omega^2 A^2/2) (ka)^4 (4/9 + 1/3) \quad (2)$$

where $4/9$ comes from the monopole and $1/3$ comes from the dipole.

If we divide by the first factor, which is the sphere cross-section, we obtain *the scattering cross-section*.

As for acoustic radiation, monopole and dipole analysis is very useful to simulate scattering if perturbation in the flow can be considered as rigid.

Whatever the different components of the scattered energy are, it is important to notice a strong dependence on wave number at power 4, which means that small wavelengths are more scattered by small obstacles than higher wavelengths.

Lambert's law for a scattering surface

Lambert's law comes from optics and is deduced from observation of scattering by surfaces. It shows that scattering signal is directly proportional to the cosine of the angle θ between the observer's line of sight and the surface normal. The law is also known as the cosine emission law, or Lambert's emission law. In this case, one speaks of *Lambertian* surfaces. It is named after Johann Heinrich Lambert, from his *Photometria*, published in 1760.

This type of scattering is taken into account particularly for analyzing the ocean; for example, scattering from the bottom of the sea, see Urlick (1976).

Let us imagine an acoustic wave intercepting the sea bottom at an incidence θ relative to the bottom.

The surface is supposed to be rough; but at a first approximation, we can think of considering the incidence and the horizontal dimension of the surface such that Snell-Descartes law can be applied.

In fact, the dimensions of surface elements relative to acoustic wavelengths can play a fundamental role, and one finds scattering radiation at angle φ relative to the surface such that if I_{inc} is the incident wave intensity, and dS a surface element, the scattered intensity I_{scatt} is found to be

$$I_{\text{scatt}} = \mu I_{\text{inc}} \sin(\theta) \sin(\varphi) dS \quad (3)$$

where μ is a coefficient of proportionality representative of the surface.

Therefore, we can analyze the relative intensity in dB:

$$P = 10 \log(I_{\text{scatt}}/I_{\text{inc}}) \\ = 10 \log(\mu) + 10 \log(\sin(\theta) \sin(\varphi)) \quad (4)$$

which gives for backward scattering ($\varphi = \pi - \theta$).

$$Pb = 10 \log(\mu) + 10 \log(\sin(\theta)^2) \quad (5)$$

This shows a variation of intensity proportional to the square of the grazing angle sinus.

In a case where transmission across the ocean bottom would be null, it is found by integration that the backscattering strength would be -5 dB.

In a same way, for a bottom transmitting no energy to the medium below, Marsh (1964) has proposed an alternative formula:

$$P = 10 \log[(\sin \theta)^4 k^3 A^2(K) / \pi \cos \theta] \quad (6)$$

where K is a wave number characteristic of the bottom sea roughness, and $A^2(K)$ the bottom-roughness power spectral density.

This expression shows that a rough bottom acts as described by Rayleigh (1878), and it is found that the roughness elements, which contribute to scatter, are related to the incident wave number and the angle θ by the relation $K = 2k \cos \theta$, which indeed has the same behavior as what is called *Bragg scattering* (from the physician Sir William Henry Bragg in 1912, in diffraction of X-rays and neutrons in crystals): enhanced backscatter due to coherent combination of signals reflected from a rough surface having features with periodic distribution in the direction of wave propagation, and whose spacing is equal to half of the wavelength as projected onto the surface.

Of course, same relationships for ocean surfaces are used for acoustic scatter at soil surfaces. The same methods used for acoustics are applied more generally to electromagnetic and optical waves (from radar and lidar) to get information on surface roughness.

The question of scatter by surface is, however, more complicated since, generally, one part of incident wave

is transmitted across the surface, and the transmission depends on surface and subsurface properties; see Urlick (1976).

Acoustic turbulent scattering

The theory of sound scattering from locally isotropic and homogeneous turbulence using the *Born approximation* (use of the incident wave in a homogeneous medium to compute the scattered wave) has benefitted from the conjunction between several researches and discoveries in the recent years: discovery of small-scale turbulence statistical behavior by Kolmogorov (1941), scattering experimental validation by Kallistratova (1959), Monin's (1962) calculation of turbulence spectra for velocity and temperature in the scattering cross-section, and some unification of the question of scattering of waves (electromagnetic and acoustical) in a turbulent medium by Batchelor (1957) and Tatarskii (1961).

The acoustic cross-section of acoustic waves in a turbulent, isotropic, and small-scale medium is

$$\eta(\theta s) = 1/8k^4 \cos^2(\theta s) \\ \times \left[\Phi_T(ks)/T_0^2 + \cos(\theta s/2)^2 E(ks)/\pi c^2 ks^2 \right] \quad (7)$$

described in Batchelor (1957) and Tatarskii (1961, 1971), where η is the differential acoustic cross-section or "reflectivity factor," θs is the angle between scattering and transmitter beam axis, k is the incident acoustic wave number, ks the scattering wave number $= 2k \sin(\theta s/2)$ corresponding to Bragg scattering, T_0 is the mean local temperature in $^\circ\text{K}$, and c is the sound speed.

An inaccuracy in the formula for the dependence of scattering was corrected in Tatarskii (1971), due to the results of Kallistratova (1959, 1960); see Kallistratova (2002).

$\Phi_T(ks)$ and $E(ks)$ are, respectively, the isotropic three-dimensional spectral densities for temperature (for a dry atmosphere) and for turbulent kinetic energy.

Let us remark that if transmitter and receiver are at the same location (monostatic system of observation), only temperature fluctuation spectra contribute to scattering, and if one considers scattering at 90° angle, no scatter occurs in this direction. In the general case, the two spectra contribute to scatter, and we can say, in this case, that we are using a bistatic configuration of observation, transmitting in one direction and receiving in another.

This formulation of scatter of sound by turbulence has been at the origin of the first Sodar by McAllister (1968). An historical and physical discussion of Sodar development can be found in Brown and Hall (1978), and in Neff and Coulter (1986).

Important points to be added correspond to small-scale turbulence definition and what scales of turbulence are responsible for backscatter are responsible:

1. Turbulence scales correspond to the so-called inertial range of turbulence, which for the atmospheric boundary layer typically corresponds to scales ranging from centimeters up to several meters.
2. Acoustic scales responsible for scatter are related to resonant scattered waves satisfying the Bragg relationship, $\lambda = 2 \lambda_s$.

Conclusion

Scattering is a very important process that takes place in the presence of rigid deformations of small scales in the propagation medium, or in the case of surface irregularities and of random perturbations in the medium. Remote sensing of atmosphere and ocean surfaces, therefore, needs a careful understanding and analysis of both medium properties and scattering processes.

Bibliography

- Batchelor, G. K., 1957. Wave scattering due to turbulence. In Sherman, F. S. (eds.), *Symposium on Naval Hydrodynamics*, NASNRC publication, N°515, National Research Council, Washington DC, Vol. 409, p. 430.
- Brown, E. H., and Hall, F. F., 1978. Advances in atmospheric acoustics. *Reviews of Geophysics and Space Physics*, **16**(171), 180. <http://mysite.du.edu/~calvert/waves/radiate.htm>
- Kallistratova, M. A., 1959. Procedure for investigating sound scattering in the atmosphere. *Soviet Physics-Acoustics*, **5**(512), 514. English translation.
- Kallistratova, M. A., 1962. Experimental studies of sound wave scattering in the atmosphere. In *Proceedings of the Third International Congress on Acoustics*. Amsterdam: EAA, ASA, DEGA, Vol. 7, p. 10.
- Kallistratova, M. A., 2002. Acoustic waves in the turbulent atmosphere: a review. *Journal of Atmospheric and Oceanic Technology*, **1139**, 1150.
- Kolmogorov, A. N., 1941. Dissipation of energy in locally isotropic turbulence. *Proceeding of USSR Academy of Sciences*, **32**, 16–18. English translation in *Proceedings of Royal Society of London, Series A* 434, 15 (1991).
- Lord Rayleigh, J. W. S., 1878. *Theory of Sound*. London: Macmillan (1st Ed., 1878, 2nd Ed., 1896). See recent Dover edition, 480 pp.
- Lord Rayleigh, J. W. S., 1945. *Theory of Sound*. New York: Dover, p. 480.
- Marsh, H. W., 1964. Sea surface statistics deduced from underwater sound measurement. *Annals of the New York Academy of Sciences*, **118**(2), 137–145, doi:10.1111/j.1749-6632.1964.tb33977.x.
- McAllister, L. G., 1968. Acoustic sounding of the lower atmosphere. *Journal of Atmospheric and Solar-Terrestrial Physics*, **30**, 1439–1440.
- Monin, A. S., 1962. Characteristics of the scattering of sound in a turbulent atmosphere. *Soviet Physics-Acoustics*, **7**, 370–373. English translation.
- Neff, W. D., and Coulter, R., 1986. Acoustic remote sensing. In Lenschow, D. (ed.), *Probing the Atmospheric Boundary Layer*. London: American Meteorological Society, Vol. 201, p. 242.
- Tatarskii, V. I., 1961. *Wave Propagation in a Turbulent Medium* (trans: Silverman, R.A.). New-York: Mac Graw Hill, 285 pp.
- Tatarskii, V. I., 1971. The effects of the turbulent atmosphere on wave propagation. In *Israel Program for Scientific Translations*,

Jerusalem, NTIS TT 68–50464, 472 pp. (Available from the National Technical Information Service, 5285 Port Royal Rd., Springfield, VA 22161).

Urick, R. J., 1976. *Principles of Underwater Sound*. New York: McGraw-Hill, 423 pp.

AEROSOLS

Ralph Kahn
NASA Goddard Space Flight Center
Greenbelt, MD, USA

Synonyms

Airborne particles; Dust; Particles; Particulate matter; PM; Smoke

Definition

Aerosols are solid or liquid particles suspended in the air, typically smaller in size than a twentieth the thickness of a human hair. There is a subtlety: In traditional aerosol science, “aerosol” refers to the particles and the medium in which they are suspended, whereas in remote sensing, the term often refers to just the particles. We use here the latter definition.

Introduction

Aerosols are of interest due to their impact on climate and health, as well as the role they might play in transporting nutrients and even disease vectors on planetary scales. Particles ranging in diameter from about 0.05 to 10 μm are studied most commonly, as they dominate aerosol direct interaction with sunlight, and are also thought to make up the majority of the aerosol mass. The particles are produced naturally by forest and grassland fires, volcanoes, desert winds, breaking waves, and emissions from living vegetation. Human activities, such as fossil fuel and agricultural burning and altering natural land surface cover, are estimated to contribute about 10 % to the global aerosol load, though these tend to concentrate near population centers where they can have both acute and long-term health consequences.

Aerosols are especially challenging to study because they originate from many, diverse sources and exhibit an enormous range of chemical compositions and physical properties. Unlike long-lived atmospheric gases, airborne particles are typically removed from the troposphere by precipitation or gravitational settling within about a week, so aerosol amount and type vary dramatically on many spatial and temporal scales. For this reason, the frequent global coverage

provided by space-based remote sensing instruments has played a central role in the study of aerosols.

Aerosol remote sensing: the first global observations

The aerosol parameter most commonly derived from remote sensing data is aerosol optical depth. It is a measure of aerosol amount based on the fraction of incident light that is either scattered or absorbed by the particles. Formally, aerosol optical depth is a dimensionless quantity, the product of the particle number concentration, the particle-average extinction cross section (which accounts for particle scattering + absorption), and the path length through the atmosphere. It is usually measured along a vertical path. Particle physical and chemical properties, such as size, brightness, and composition, must also be either measured or assumed, to assess aerosol impact on climate and health. Measuring particle properties remotely remains a major challenge to the field.

Since late 1978, the Advanced Very High Resolution Radiometer (AVHRR) imagers have been collecting daily, global, multispectral data, from polar orbit. These instruments were designed primarily to observe Earth's surface, and data analysis typically included an "atmospheric correction" aimed at eliminating surface feature blurring caused by the intervening gas and particles. However, over-ocean, total-column aerosol optical depth was also deduced, initially from single-channel observations, assuming a completely dark ocean surface in the red band (0.63 μm wavelength) and medium-sized, purely scattering particles.

Global, seasonal, and shorter-term aerosol distributions have been mapped. By associating observed particle concentrations with deserts, wildfire regions, and high-population areas on nearby land, dust, smoke, and aerosol pollution plumes were identified. The AVHRR maps showed that Saharan dust is routinely carried across the Atlantic Ocean and deposited in the Caribbean and, more generally, demonstrated the degree to which aerosols are carried long distances. The observations led to interest in the possible impact of aerosols on the energy balance and hence the climate of the planet, and the role desert dust might play in fertilizing iron-poor remote oceans and the Amazon basin.

Among the limitations of this early work was the inability to separate surface from atmospheric contributions to the observed scene brightness, which precluded retrievals over brighter coastal regions and most land. There was also a lack of direct information about particle properties in these retrievals, and an absence of constraints on vertical distribution, both of which are important for calculating aerosol impacts on climate and health. In addition, it was difficult to distinguish cloud from aerosol signals and, more generally, to assess the accuracy of the reported values. Enormous efforts and considerable

progress were made in each of these areas over the subsequent 25 years.

One way to avoid significant surface contributions to the observed signal is to view the planet edge on, as was done by the Stratospheric Aerosol and Gas Experiment (SAGE) instruments, beginning in 1979. Observing the sun through the long slant path of the atmosphere as the satellite crossed to the night side of Earth, SAGE produced upper atmosphere vertical soundings, which proved immensely effective for monitoring the sulfate aerosols produced in the stratosphere by the Mt. Pinatubo eruption in 1991. But with no more than two occultations per orbit (about 30 per day), and limited ability to sound even as far down as the upper troposphere, other approaches were needed to address key aerosol-related climate and health questions.

Due to the intense scattering of ultraviolet (UV) light by atmospheric molecules, the surface is obscured when viewed from space in the wavelength range 330–380 nm. In the late 1990s, it was realized that the Total Ozone Mapping Spectrometers (TOMS), versions of which had already been orbiting for nearly 20 years, contained the spectral channels needed to retrieve aerosol amount over land as well as water, based on their ability to absorb the upwelling background UV light. This resulted in the Aerosol Index, a qualitative measure of UV-absorbing aerosols such as dust and smoke. The retrieval has limited sensitivity to near-surface aerosols and depends sensitively on the elevation of the particles and their optical properties, but the maps provided the first comprehensive, long-term record of aerosol source regions and overland transports.

One of the first and most widely used aerosol remote sensing techniques is surface-based sun photometry, which involves measuring the varying intensity of the solar disk as the sun changes elevation in the sky. The method predates satellite observations and relies on observing a systematic increase in atmospheric opacity (and the corresponding decrease in solar brightness) as the sun is viewed through longer atmospheric slant paths. Assuming aerosol horizontal homogeneity, column-integrated aerosol optical depth is retrieved, and if this is done at multiple wavelengths, some information about particle size can also be derived.

Beginning in 1991, with the European Space Agency (ESA) two-view-angle Along Track Scanning Radiometer (ATSR) series of imagers exploited the geometrically based approach from space. Unlike sun photometry, the satellite technique measures light scattered by the scene below, so additional assumptions about aerosol and surface optical properties are required to retrieve column-integrated aerosol optical depth. But the atmospheric contribution to the signal still increases systematically, relative to that of the surface, as the slant path increases, making surface-atmosphere separation possible. The steeper slant paths also improve sensitivity to thinner aerosol layers.

In 1996, the first of the French Space Agency (CNES) POLarization and Directionality of Earth's Reflectance (POLDER) imagers began collecting multi-angle, multi-spectral polarization data from orbit. The polarization effects of many types of land surfaces are fairly independent of wavelength, making it possible to separate the more constant surface polarization contribution to the satellite signal from the spectrally varying atmospheric contribution. Aerosols are sometimes divided into two groups, depending on whether their effective diameter is greater or less than a certain size, usually taken to be around 1 μm for satellite observations and 2.5 μm for many health and direct sampling applications. The majority of smoke and aerosol pollution particles, the products of combustion and chemical processing, tend to fall into the smaller or "fine" mode. Mechanically produced desert dust and sea salt particles tend to be weighted toward the "coarse" mode. From its combination of optical measurements, POLDER maps column-integrated fine-mode and total aerosol optical depth over water, as well as fine-mode optical depth overland, and more advanced retrieval algorithms for POLDER data are under development.

Second-generation global measurements

The NASA Earth Observing System's (EOS) Terra satellite, launched in late 1999 into a sun-synchronous orbit crossing the equator at about 10:30 local time, carries two instruments designed in part to make detailed aerosol measurements: the MODerate resolution Imaging Spectroradiometer (MODIS) and the Multi-angle Imaging SpectroRadiometer (MISR). MODIS follows the multi-spectral approach of the AVHRR instruments, but with higher spatial resolution (a maximum of 250 m, compared with up to 1 km for AVHRR), 36 spectral channels, and much higher radiometric calibration accuracy and stability. Total-column aerosol optical depth over global water as well as darker land surfaces is produced routinely every 2 days, along with fine-mode fraction over ocean. Thermal infrared channels are used to detect fires, which can help identify smoke plumes. Efforts are being made to extend the interpretation of MODIS data, for example, by using blue-band data to retrieve aerosol optical depth over bright surfaces (following the TOMS approach) and by developing semiempirical ways of deducing aerosol type from the combination of geographic location, season, and fine-mode fraction. A second copy of MODIS, flying on the EOS Aqua satellite, provides observations similar to MODIS/Terra, at about 1:30 equator crossing time.

MISR complements MODIS, acquiring four-channel, near-simultaneous multispectral views of Earth at nine angles and spatial sampling up to 275 m/pixel. Having a narrower swath than MODIS, MISR takes about a week to image the entire planet. With the multi-angle coverage, aerosol optical depth as well as surface reflectance are retrieved, even over bright desert surfaces. The nine views also sample light scattered in different

directions, which yields some information about column-averaged particle size, shape, and brightness under good retrieval conditions; taken together, the retrieved information makes it possible to map aerosol air mass types, though not detailed particle microphysical properties. From the stereo viewing, MISR also derives the heights of clouds and aerosol plumes near their sources. The injection heights of wildfire smoke, volcanic effluent, and desert dust, produced by stereo imaging, are key quantities used in computer-based aerosol transport simulation models to predict plume evolution and aerosol climate impact.

The most accurate information about aerosol vertical distribution is obtained from lidars, which, unlike passive imagers that collect scattered sunlight, send out their own laser beam as the illumination source. Their ability to detect multiple, very thin aerosol layers, day and night, makes it possible for them to create a global, climatological picture of transported aerosols. But the lidar swath is the width of the laser beam, less than a 100 m in size, so coverage of specific sites or events is serendipitous. The technique was demonstrated in 1994, with the Lidar In-Space Technology Experiment (LITE) aboard NASA's Space Shuttle Discovery, and has been followed by the Geoscience Laser Altimeter System (GLAS) and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) instruments on polar-orbiting EOS satellites. CALIPSO flies in a constellation of satellites called the "A-Train," aimed at making complementary, near-coincident measurements. It includes Aqua carrying one MODIS instrument as well as PARASOL, and the Aura satellite bearing updated versions of POLDER and TOMS, respectively.

The second-generation measurements also represent a new era for assessing the accuracy of satellite aerosol retrieval results. Surface-based networks of autonomous sun photometers, such as the several-hundred-instrument AEROSOL ROBOTIC NETWORK (AERONET) federation, began producing very accurate, uniformly processed time series of aerosol column amount. The results are compared statistically and event by event with the satellite retrievals. By scanning across the sky in addition to making direct sun measurements, the surface stations collect the data needed to derive column-average particle size and brightness as well. Regional surface-based networks of radiometers also contribute to satellite retrieval validation, as well as surface lidar networks such as the NASA's Micro Pulse Lidar Network (MPLNET) and European Aerosol Research Lidar Network (EARLINet).

Despite enormous advances in space- and ground-based aerosol remote sensing, some key measurements elude these techniques. Currently, detailed knowledge of particle composition, brightness, shape, and ability to absorb water is obtained only by collecting samples of the particles themselves. For near-surface aerosols such as urban pollution particles, surface stations are usually deployed, equipped to measure particle size and mass and to acquire samples for chemical analysis. But for more complete

atmospheric characterization, intensive field campaigns are required, involving aircraft, satellites, and ground stations making coordinated observations, often with the help of predictions from aerosol transport models.

Aerosol environmental effects

Most aerosols scatter more than 90 % of the visible light falling upon them, whereas nearly 70 % of Earth's surface is dark water. So overall, aerosols tend to cool the planet by making it more reflecting than it would otherwise be. Climate models play a central role in assessing the magnitude of this effect, as polar-orbiting satellites provide only snapshots of aerosol optical depth and limited information about their brightness, and aerosol data from geostationary satellites, which can make continuous observations of the sub-spacecraft region, are currently only qualitative. For radiative forcing calculations, models simulate the full diurnal cycle, along with uniform, global fields of aerosol amount, brightness, and vertical distribution. Remote sensing observations provide constraints on these models.

Current assessments suggest that the globally averaged aerosol direct radiative effect amounts to a cooling of about 0.1–0.9 W/m², compared to the global warming by carbon dioxide of about 1.66 W/m² (IPCC, 2007). Although the aerosol cooling might offset some greenhouse warming, aerosols, unlike long-lived greenhouse gases, are not distributed uniformly, so their regional effects are far more significant than the global mean. Preliminary assessments of observed trends in aerosol optical depth suggest that since the mid-1990s, aerosol particle pollution has decreased over Europe and eastern North America, whereas it has increased over east and south Asia, and on average, the atmospheric concentration of low-latitude smoke particles has also increased.

In addition to the direct effect they have on sunlight, aerosols play a role in the formation of clouds. Collecting the water molecules needed to make cloud droplets is accomplished with the help of aerosols, called cloud condensation nuclei (CCN). The concentration of CCN in the droplet-formation regions of clouds mediates the number of droplets that form. If a fixed amount of cloud water is divided into more droplets of smaller size, the overall reflectivity of the cloud will be greater. The “first indirect effect” of aerosols on clouds refers to the way increased CCN can lead to increased cloud reflectivity. As smaller-sized droplets are less likely to coalesce into raindrops and precipitate, smaller droplets can also increase cloud lifetime, increasing global reflectivity and slowing the cycling of water through the atmosphere, a process often called the second indirect effect.

Aerosols can weaken or strengthen clouds in other ways as well. If dark aerosols such as soot or smoke are present, they can absorb sunlight and evaporate cloud droplets. Aerosols dissolved in cloud droplets may postpone freezing as droplets are carried aloft in the cores of some types of clouds, invigorating their development. Phytoplankton in ocean surface water might actively

regulate cloud amount by modulating their output of gaseous sulfur compounds, which in turn form sulfate aerosols in the atmosphere and serve as CCN in aerosol-poor skies over remote oceans. It is difficult to test these hypotheses on large scales with currently available measurements, and therefore to assess their environmental impacts, but observations indicate that such mechanisms must at least be considered as part of the global picture.

The magnitude of aerosol indirect effects on clouds is much less certain than that of aerosol direct radiative effects, in part because detailed particle size and composition, which determine their ability to absorb water, cannot be measured remotely with sufficient accuracy. In addition, most CCN are smaller than 0.05 µm, too small to be distinguished from atmospheric gas molecules by remote sensing techniques. The current consensus model-based estimate for global-average indirect effects is cooling of 0.3–1.8 W/m², but with low confidence (IPCC, 2007). Narrowing the uncertainties is likely to require a combination of detailed chemistry and microphysics from aircraft measurements, satellite observations to provide broad statistical sampling, and modeling to generate an overall result.

To date, remote sensing has contributed to our knowledge of aerosol impacts on human health only qualitatively, identifying, with the help of aerosol transport models or lidar profiles, regions where near-surface particle concentrations are especially high. Surface stations are typically more effective in isolating the near-surface aerosols that matter most for these studies, and direct samples are required to provide detailed particle size and chemical composition. However, the spatial coverage of surface stations is very limited. As with climate effects, the future seems to point toward combining satellite and suborbital measurements with models.

Summary

Our appreciation of aerosols' role in climate change has grown over the past 25 years, in part due to the contributions made by remote sensing. First estimates of the impacts transported aerosols have on the atmospheric energy balance, on clouds and the hydrological cycle, on larger-scale atmospheric circulation, and on human health have been made. An understanding has developed for the need to combine detailed physical and chemical measurements from aircraft and ground stations and extensive constraints on aerosol optical depth, type, and vertical distribution from satellites, with numerical models that can simulate present and predict future conditions.

However, much remains to be done. For planning purposes, the accuracy of measurements needed to assess aerosol direct radiative effect must be improved, and uncertainties in aerosol indirect effects on clouds must be reduced. Techniques for systematically constraining models with satellite and suborbital data need to be developed, both to test model parameterizations of aerosol sources, cloud processes, etc., and to assess the

uncertainties in the resulting simulations. Based on past experience, this can be achieved, provided we continue to develop and deploy the instruments, improve the models, and maintain the research community, which have carried the field to this point.

Bibliography

- Charlson, R. J., Lovelock, J. E., Andreae, M. O., and Warren, S. G., 1987. Ocean phytoplankton, atmospheric sulphur, cloud albedo, and climate. *Nature*, **326**, 655–661.
- Haywood, J., and Boucher, O., 2000. Estimates of the direct and indirect radiative forcing due to tropospheric aerosols: a review. *Reviews of Geophysics*, **34**, 513–543.
- IPCC, 2007. The physical science basis. In Solomon, S., Qin, D., Manning, H., Chen, Z., Marquis, M., Averyt, K., Tignor, M., Miller, H. (eds.), *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Kahn, R. A., 2012. Reducing the uncertainties in direct aerosol radiative forcing. *Surveys in Geophysics*, **33**, 701–721, doi:10.1007/s10712-011-9153-z.
- Kaufman, Y. J., Tanre, D., and Boucher, O., 2002. Satellite view of aerosols in the climate system. *Nature*, **419**, 215–223, doi:10.1038/nature01091.
- Ramanathan, V., Crutzen, P. J., Kiehl, J. T., and Rosenfeld, D., 2001. Aerosols, climate, and the hydrological cycle. *Science*, **294**, 2119–2124.
- Seinfeld, J. H., and Pandis, S. N., 2006. *Atmospheric Chemistry and Physics*, 2nd edn. New York: Wiley, p. 1203.

AGRICULTURAL EXPANSION AND ABANDONMENT

Jiaguo Qi

Department of Geography/CGCEO, Michigan State University, East Lansing, MI, USA

Synonyms

Agricultural abandonment; Agricultural expansion; Agricultural land; Change detection; Crop phenology; Mapping agricultural lands

Definition

Agricultural land. Agricultural land may be defined broadly as land used primarily for the production of food and fiber, including croplands, pastures, orchards, groves, vineyards, nurseries, ornamental horticultural areas, confined feeding operations, and other agricultural applications.

Agricultural expansion. The conversion of nonagricultural land use to agricultural use.

Agricultural abandonment. The conversion of agricultural land use to nonagricultural use.

Change detection. The process of identifying and documenting changes in land use over time.

Crop phenology. The study of changes in plant physiology and crop growth stages as influenced by environmental

and anthropogenic drivers, including, for example, seasonal variations in temperature and precipitation.

Mapping agricultural lands. Interpreting remotely sensed imagery to categorize agricultural lands.

Introduction

Agricultural lands are by far the largest land areas that have been intensively managed for food and fiber production. Increases in food and fiber demand have driven significant agricultural expansion over the past decades. The rate and magnitude of agricultural expansion are ever increasing, especially under the recent pressure to develop sustainable energy sources such as biofuels from agricultural crops. Agricultural expansion is the *conversion* of nonagricultural lands to agricultural lands, for example, conversion of forested lands to crops, wetlands to rice fields or fishponds, or natural grasslands to pastures.

At the same time, agricultural lands in many places have been abandoned as traditional agricultural operation has become either unprofitable or impracticable. Agricultural abandonment can be a result of various factors ranging from natural to human. The natural factors may include climate change, such as decline in rainfall or the occurrence of extended periods of drought or flood, which makes traditional agricultural operation practically impossible. In places where agricultural operations rely heavily or exclusively on irrigation, agricultural abandonment may occur due to salinization or reduction in ground or surface water supplies. In some cases, agricultural abandonment happens solely due to economic pressures such as the conversion of traditional agricultural lands into suburban real estate developments near major cities around the world.

Agricultural expansion or abandonment has significant consequences for the economy and food security (Alexandratos, 1999; De La Torre Ugarte et al., 2003; Walsh et al., 2003) as agriculture is the backbone of food supplies. However, agricultural expansion can cause detrimental effects on nonagricultural terrestrial and aquatic ecosystems (Tilman et al., 2001). These environmental consequences include increased soil erosion, nutrient leaching that pollutes aquatic environments and degrades water quality, reduced water availability for other ecological uses, and changes in biogeochemical processes that impact greenhouse gas emissions and soil carbon sequestration.

Whether through expansion or abandonment, agricultural land uses are being altered, and these changes will affect food security, the environment, and future sustainability. The role of remote sensing in agricultural expansion or abandonment is to provide detailed assessments of the spatial and temporal changes of agricultural and nonagricultural lands. Effective monitoring of agricultural expansion and/or abandonment provides timely information for decision makers, economists, and environmental scientists to objectively assess future food security and environmental consequences.

Remote sensing of agricultural expansion and abandonment

Remote sensing has been used as a cost-effective tool to map the spatial extent of agricultural lands (Hogg, 1986; Hall and Badhwar, 1987; Congalton et al., 1998). With multi-temporal remote sensing, agricultural expansion or abandonment can effectively be mapped through change detection analysis. As such, remote sensing of agricultural expansion or abandonment requires a minimum of two images acquired on different dates covering the period for documenting changes in agricultural lands. Two common approaches are utilized for this application: (1) comparison of separately classified remote sensing imagery from two or more different dates and (2) direct, single classification of multi-date remote sensing imagery. These change detection approaches are used to quantify agricultural expansion and abandonment.

Mapping agricultural lands

Agricultural mapping using remotely sensed imagery is a common practice where observed spectral radiance or reflectance is classified into different land use categories. Techniques often employed include supervised and unsupervised classification using various cluster analyses (Haralick et al., 1970; Wharton, 1982; Singh, 1989; Congalton et al., 1998; Steele, 2000; Lucas et al., 2007; Karjalainen et al., 2008). The fundamentals of these classification techniques rely on the fact that different land cover types or crops have distinct spectral reflectance properties and patterns. By analyzing their spectral properties or spatial patterns, similar agricultural lands or crops can be grouped together as a single class (Haralick et al., 1973; Bischof et al., 1992; Vailaya et al., 1998; Benediktsson and Kanellopoulos, 1999).

A recent improvement in agricultural land mapping involves the use of phenological properties of crops. Crop phenological information can be derived from repeated satellite observations, such as the normalized difference vegetation index, that track the changes in plant growth and physiological stage. Because different crops have different phenological properties, such as number of days to reach maximum growth, they can be effectively identified or mapped by analyzing the time series of remote sensing observations (Haralick et al., 1980; Jakubauskas et al., 2002; Wang and Tenhunen, 2004; Chang et al., 2007; Wardlow et al., 2007; Busetto et al., 2008). Use of phenology for mapping crop types is a relatively new application due to the recent availability of frequent, repeated remote sensing observations.

Change detection and assessment

Commonly used techniques of change detection include (a) multi-temporal composite image analysis, (b) image algebra (e.g., band differencing and band ratioing), (c) post-classification comparison, (d) binary masking, (e) spectral change vector analysis, (f) cross-correlation, (g) visual on-screen digitization, and (h) knowledge-based

vision systems (e.g., Singh, 1989; Congalton et al., 1998; Jensen, 2005). For quantitative assessment of agricultural expansion or abandonment, often the post-classification method is used. It requires co-registration and classification of each remotely sensed image. The two classified maps are then compared on a pixel-by-pixel basis using a change detection matrix. This matrix provides quantitative “from-to” land use change information, thereby documenting the total amount of agricultural lands that have either expanded or been abandoned.

Care must be taken for accurate agricultural change detection and assessment. Image processing/interpretation must account for phenological differences of crop systems and differences in spatial resolution, reduce radiometric artifacts from external factors such as atmospheric and substrate conditions, and ensure accurate co-registration of imagery. The accuracy of change detection depends on the effectiveness of all these image processing procedures. Therefore, high-quality image processing and classification is imperative to accurately assess agricultural expansion or abandonment.

Current obstacles to mapping agricultural expansion/abandonment

Agricultural lands are very dynamic, changing rapidly with time. Remote sensing imagery provides information only about the physical attributes of land cover and not on the “purpose” of the lands, or the intentions of the people who manage the lands. For example, forest is a cover type and can be sensed remotely, but whether the forested land is intended for timber harvesting or recreational uses needs to be determined by the owner. A major obstacle in mapping agricultural land use changes is that while remote sensing can identify land cover changes, these changes may not be the result of agricultural land use changes. Successful mapping of agricultural expansion and abandonment is the differentiation of temporary land cover changes (e.g., short-term fallow) from long-term land use changes. Since temporary land cover changes can have the same or similar spectral properties, mapping agricultural expansion or abandonment requires multiple year observations and sufficient ancillary information in order to identify observed changes in land cover as land use changes.

Another obstacle is related to the nature of agricultural lands themselves. Many agricultural crops have similar spectral properties, phenological development, and spatial patterns as nonagricultural cover types. There may be no spectral or spatial pattern differences across these lands even though their uses are different. This can make accurate imagery classification of agricultural expansion/abandonment mapping very difficult.

Conclusion

Agricultural expansion and abandonment is widespread globally and one of the most important forms of land use change. Accurate assessment and monitoring of the rate and spatial extent of expansion and abandonment provides

vital information about the location and availability of agricultural lands for economists, environmentalists, biologists, ecologists, resource managers, and policy makers. Remote sensing provides a cost-effective and repeatable means of mapping and assessing the spatial pattern and rate of change in agricultural land use. With the capabilities of future advanced earth observation systems, such as the *Global Land Observation System* (GLOS) and the *Global Earth Observation System of Systems* (GEOSS), it is foreseeable that timely information about agricultural expansion and abandonment will become available to a much broader user community.

Bibliography

- Alexandratos, N., 1999. World food and agriculture: outlook for the medium and longer term. *PNAS: Proceedings of the National Academy of Sciences*, **96**(11), 5908–5914.
- Benediktsson, J. A., and Kanellopoulos, I., 1999. Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transactions on Geoscience and Remote Sensing*, **37**(3), 1367–1377.
- Bischof, H., Schneider, W., and Pinz, A. J., 1992. Multispectral classification of landsat-images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, **30**(3), 482–490.
- Busetto, L., Meroni, M., and Colombo, R., 2008. Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series. *Remote Sensing of Environment*, **112**, 118–131.
- Chang, J., Hansen, M. C., Pittman, K., Carroll, M., and DiMiceli, C., 2007. Corn and soybean mapping in the United States using MODIS time-series data sets. *Agronomy Journal*, **99**, 1654–1664.
- Congalton, R. G., Balogh, M., Bell, C., Green, K., Milliken, J. A., and Ottman, R., 1998. Mapping and monitoring agricultural crops and other land cover in the Lower Colorado River Basin. *Photogrammetric Engineering and Remote Sensing*, **64**(11), 1075–1118.
- De La Torre Ugarte, D., Walsh, M., Shapouri, H., and Slinksy, S., 2003. *The economic impacts of bioenergy crop production on U.S. agriculture*. U.S. Department of Agriculture, Office of the Chief Economist, Agricultural Economic Report 816.
- Hall, F. G., and Badhwar, G. D., 1987. Signature-extendable technology: global space-based crop recognition. *IEEE Transactions on Geoscience and Remote Sensing*, **25**(1), 93–103.
- Haralick, R. M., Caspell, F., and Simonett, D. S., 1970. Using radar imagery for crop discrimination: a statistical and conditional probability study. *Remote Sensing of Environment*, **1**, 131–142.
- Haralick, R. M., Shanmugam, K., and Dinstein, I., 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, **3**(6), 610–621.
- Haralick, R. M., Hlavka, C. A., Yokoyama, R., and Carlyle, S. M., 1980. Spectral-temporal classification using vegetation phenology. *IEEE Transactions on Geoscience and Remote Sensing*, **18**(2), 167–174.
- Hogg, H., 1986. Agriculture and resource inventory survey through aerospace remote sensing (AgRISTARS). *IEEE Transactions on Geoscience and Remote Sensing*, **24**(1), 185.
- Jakubauskas, M. E., Legates, D. R., and Kastens, J. H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and Electronics in Agriculture*, **37**, 127–139.
- Jensen, J. R., 2005. Digital change detection. In Jensen, J. R. (ed.), *Introductory Digital Image Processing: A Remote Sensing Perspective*. Englewood Cliffs: Prentice-Hall, pp. 467–494.
- Karjalainen, M., Kaartinen, H., and Hyypä, J., 2008. Agricultural monitoring using Envisat alternating polarization SAR images. *Photogrammetric Engineering and Remote Sensing*, **74**(1), 117–128.
- Lucas, R., Rowlands, A., Brown, A., Keyworth, S., and Bunting, P., 2007. Rule-based classification of multi-temporal satellite imagery for habitat and agricultural land cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, **62**(3), 165–185.
- Singh, A., 1989. Review article digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, **10**(6), 989–1003.
- Steele, B. M., 2000. Combining multiple classifiers. An application using spatial and remotely sensed information for land cover type mapping. *Remote Sensing of Environment*, **74**(3), 545–556.
- Tilman, D., 1999. Global environmental impacts of agricultural expansion: the need for sustainable and efficient practices. *Proceedings of the National Academy of Sciences*, **96**(11), 5995–6000.
- Tilman, D., Fargione, J., Wolff, B., and d'Antonio, C., 2001. Forecasting agriculturally driven global environmental change. *Science*, **292**, 281–284.
- Vailaya, A., Jain, A., Zhang, H. J., 1998. On image classification: city versus landscape. In *IEEE Workshop on Content – Based Access of Image and Video Libraries (CBAIVL)*, Santa Barbara, CA, p. 3.
- Walsh, M., De La Torre, U. D., Shapouri, H., and Slinksy, S., 2003. The economic impacts of bioenergy crop production on U.S. Agriculture. *Environmental and Resource Economics*, **24**, 313–333.
- Wang, Q., and Tenhunen, J. D., 2004. Vegetation mapping with multitemporal NDVI in North Eastern China Transect (NECT). *International Journal of Applied Earth Observation and Geoinformation*, **6**, 17–31.
- Wardlow, B. D., Egbert, S. L., and Kastens, J. H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment*, **108**, 290–310.
- Wharton, S. W., 1982. A contextual classification method for recognizing land use patterns in high resolution remotely sensed data. *Pattern Recognition*, **15**(4), 317–324.

Cross-references

[Agriculture and Remote Sensing](#)
[Data Processing, SAR Sensors](#)
[Global Earth Observation System of Systems \(GEOSS\)](#)
[Global Land Observing System](#)
[Vegetation Phenology](#)

AGRICULTURE AND REMOTE SENSING

Jerry Hatfield¹ and Susan Moran²

¹National Laboratory for Agriculture and the Environment, Ames, IA, USA

²USDA ARS Southwest Watershed Research Center, Tucson, AZ, USA

Definition

Reflectance. Light that is returned from the surface of an object in the same wavelength that impinged on the object.

Emission. Emission of energy in wavelengths determined by the Stefan-Boltzmann relationship.

Vegetative index. Combination of wavelengths that are related to a specific canopy parameter.

Canopy parameters. Descriptions of factors that physically define the canopy, e.g., height, leaf area, biomass, yield.

Thermal index. Comparisons of canopy and air temperature that are related to crop water deficits.

Introduction

Agricultural scientists have used remote sensing for hundreds of years to observe plants to assess their vigor or stress from a multitude of factors. These original observations were not made with sensors but with the eye that determined the health of the plant. The calibration process was to compare the affected plant against a standard that the individual had observed before and deemed to be healthy. This type of analysis was possible because there was a change in the color of the leaf which the eye detected as a change from normal and thus different from what was expected. Visual observations remain a viable tool for crop assessment.

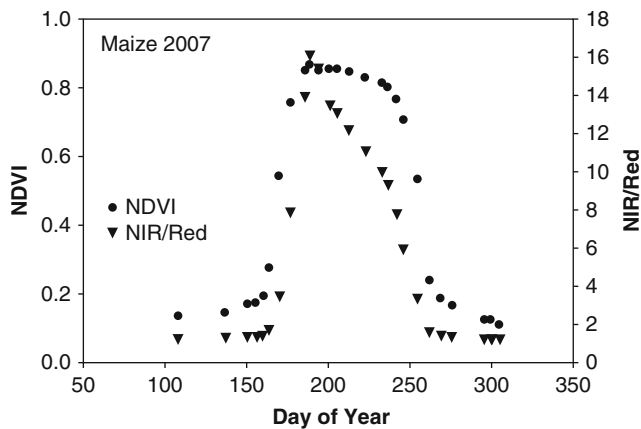
Remote sensing in agriculture is possible because of the changes in leaf reflectance throughout the visible, near-infrared, and shortwave infrared wavelengths as shown in Fig. 1. A healthy leaf has a larger reflectance in the green portion of the spectrum with smaller reflectances in the blue and red portions. Leaves are unique in that their reflectance rapidly increases near 0.7 μm and remains high throughout the near-infrared region (Fig. 1). These changes in reflectance allow the leaf to avoid the absorption of large amounts of energy that it would have to dissipate through some means of energy exchange. Light absorbed in the visible regions drives the photosynthetic process. Leaves appear green to our eyes because these wavelengths have the highest reflectance. If our eyes could “see” in the near-infrared wavelengths, leaves

would appear extremely bright because of the high reflectance. Deviations from the expected patterns provide an indication of some level of stress.

The use of spectrophotometers to measure the optical properties of leaves were presented in the early work by William Allen, David Gates, Harold Gausman, and Joseph Woolley who described the basic theory relating morphological characteristics of crop plants to their optical properties (Allen et al., 1969; Gates et al., 1965; Gausman et al., 1969; Woolley, 1971). Their efforts provided detail about high-resolution spectral signatures of natural and cultivated species and were sources of information about normal plant growth and conditions caused by nutrient deficiency, pests, and abiotic stresses (Gausman and Allen, 1973). These research findings help describe, for the first time, the reasons why the spectral signatures of leaves varied from the expected values and the information contained by examining these deviations. Through these basic studies the foundation was formed for what is applied today as remote sensing tools for agriculture. It is not possible to list all of the publications generated by these pioneers in remote sensing but their efforts help develop the basic understanding of leaf reflectance that is utilized today in remote sensing of agricultural crops.

Leaves are not the same as plant canopies and reflectance values for canopies are more complex than individual leaves. Canopies have many different morphological forms because of branch arrangement on trees and vines, leaf architecture and positioning on annual and perennial plants, whether plants are deciduous or evergreen, and how plants are arranged in different planting configurations. Observations of different plants provide an indication of how different canopies reflect light. There is a vast array of potential shapes of canopies, and light interactions with canopies are different than individual leaves. It is this interaction of light with plant canopies that has provided for the value of remote sensing of agricultural canopies and the potential value for using reflectance from different canopies to assess various canopy properties.

The thermal portion of the spectrum represents the longwave radiation and the wavelengths emitted are determined by the Stefan-Boltzmann relationship in which the wavelengths emitted are function of the fourth power of the temperature of the object. Objects on the earth’s surface emit in wavelengths between 8 and 14 μm because of their temperatures being in the range of 300 K. Remote sensing of crop surfaces using infrared thermometers has become routine and offers the potential for crop water stress detection and water management. Similar to the visible and near-infrared research, understanding the importance of thermal radiation for agricultural crops can be traced to original observations by Tanner (1963) who found that plant temperature varied from air temperature and could be measured with thermocouples attached to the leaves. Since that basic discovery, there have been continual advances in the use of thermal wavelengths in agricultural applications.



Agriculture and Remote Sensing, Figure 1 Reflectance of an individual leaf in the visible and near-infrared portions of the light spectrum.

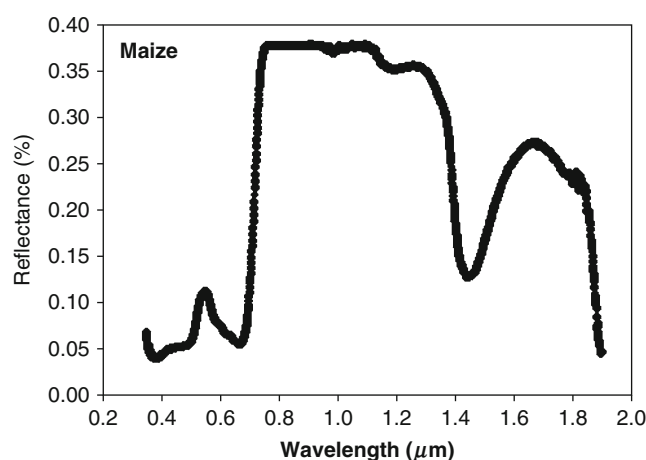
There has been considerable progress in the field of remote sensing and agriculture and the progress was recently documented in Volume 69 of *Photogrammetric Engineering and Remote Sensing* in 2003. The topics covered in this volume were sensor techniques (Barnes et al., 2003), hydrometeorological applications (Kustas et al., 2003), crop management applications (Pinter et al., 2003), crop yield assessment (Doraiswamy et al., 2003), applications to rangeland assessment and management (Hunt et al., 2003), water quality assessment (Ritchie et al., 2003), and sensor development and correction methods (Moran et al., 2003). Another recent review of the application of remote sensing methods to dryland crops was developed by Hatfield et al. (2004) and the advances of remote sensing to agronomic decisions by Hatfield et al. (2008).

Canopy parameters

Observations of canopies provide information about the typical parameters that are observed through other means. These include biomass, leaf area, ground cover, chlorophyll content of the canopies, light interception, grain yield, and crop residue on the soil surface. Each of these parameters provides critical information that can be used to determine if the progression of the crop is at a normal rate or is deviating either in a negative or positive direction away from the normal expectations. It is important to detect how quickly the crop is developing during the season and the size of the plant.

Canopies change throughout the growing season and it is that change that provides the information that can be detected from remote sensing platforms. Agricultural canopies vary in their size, row direction, canopy shape, and even color. Annual crops are different from perennial crops such as orchards or vineyards. Agricultural systems like pastures or rangelands that are permanent grasses or shrubs display large changes throughout the year. It is important to realize that each crop is different, and crop identification has been a primary goal of many remote sensing studies. These crop identification approaches have been used to develop land use maps and crop area for the earth's land area.

Canopy reflectance is different from leaves because of the soil background beneath the crop. Canopy reflectance varies throughout the year, and if we apply the spectral maps to a whole canopy, as compared to an individual leaf, remarkable differences can be observed, as shown in Fig. 2 for a corn crop. There are several important features in this figure that reveal information about the canopy. First, the reflectance of bare soil when there is no canopy is linear and exhibits little difference among individual wavelengths. Second, the variation in the reflectance values for the bare soil is caused by changes in surface roughness induced by either tillage or rainfall events. Third, there is a rapid change in the reflectance values in the visible and near-infrared portions as the canopy begins to develop and grow. Coupled with the rapid increase in the

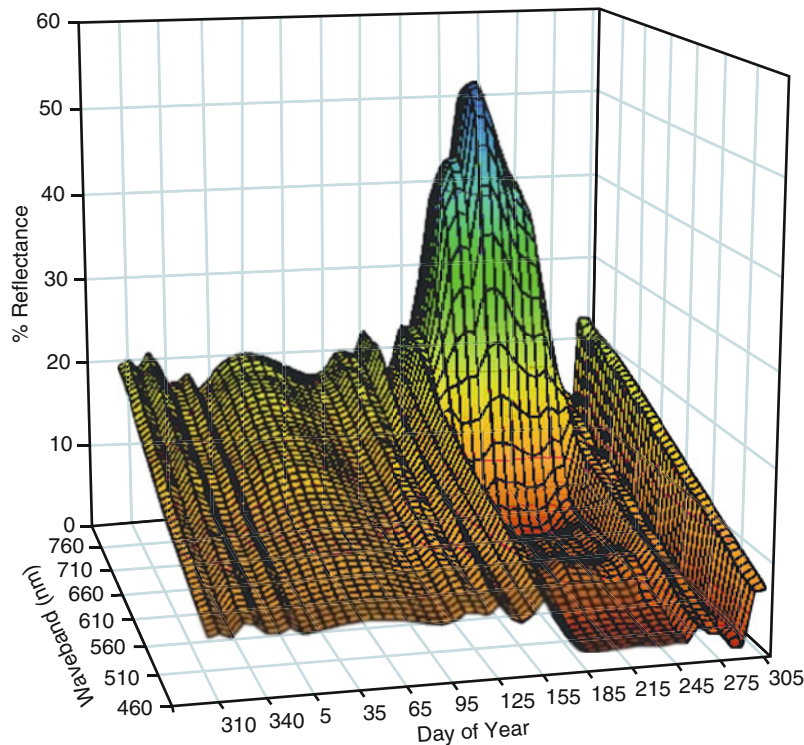


Agriculture and Remote Sensing, Figure 2 Reflectance across the visible and near-infrared portions of the spectrum for a corn canopy across the year.

near-infrared reflectance values, there is a concurrent decrease in the blue and red and an increase in the green wavelengths. Fourth, there is an equally rapid decline in the near-infrared reflectances as the canopy begins to senesce and the green leaves turn brown. There are very dynamic changes that occur throughout the growing season and these dynamics provide the potential information content from tracking the reflectance patterns of different wavelengths.

Vegetative indices

Vegetative indices (VIs) has been used as a term that describes the relationship of a combination of wavebands or wavelengths to a particular canopy parameter. These VIs range from simple ratios of wavebands to complex empirical relationships of different wavebands or even combinations of different VIs for specific purposes. These have been developed and evaluated extensively over the past 30 years which coincides with the acquisition of reflectance data from the Landsat platforms. These satellites were launched with four broad wavebands that have been extensively used in the development and application of a number of VIs. A summary of the different VIs provided by Jackson and Huete (1991) stated the purpose of VIs was to enhance the vegetation signal while minimizing the solar irradiance and soil background effects. The adjustment for soil background effects is extremely critical in using remote sensing because annual crop canopies and many perennial tree crops are planted in rows. Developing crop canopies do not completely cover or shade the soil; therefore, viewing the crop from above presents a problem in which the signal is a mixture of the bare soil reflectance values and the leaf reflectance. The canopy reflectance is a mixture of both surfaces and VIs need to be sensitive to soil background and be able to compensate for these differences.



Agriculture and Remote Sensing, Figure 3 Seasonal changes in NIR/RED reflectance and normalized difference vegetative index (NDVI) for a maize canopy in central Iowa.

The development of VIs can be traced back to Jordan (1969) who related the ratio of NIR (0.8 μm) to red (0.675 μm) reflectance (NIR/RED) to leaf area index (LAI). Tucker (1979) found it was possible to use red and near-infrared as measures of changes in vegetation over large areas and increased the confidence that VIs had an application for large-scale assessment of crop canopies. He proposed a difference vegetative index (DVI) as NIR-RED. Development of VIs either as ratios of wavebands or linear combinations of wavebands has been primarily through statistical analyses of observed crop parameters and reflectance data. The seasonal trajectory of reflectance from a crop canopy reveals the red reflectance reaching a minimum with increasing leaf area while the near-infrared (NIR) increases (Fig. 2). This causes the ratio to increase as the canopy develops and a range of NIR/RED values from about 1 to over 20. One of the major limitations in the NIR/RED ratio is the lack of sensitivity when the ground cover is low and both reflectance values are similar. Deering (1978) found this limitation could be overcome through the use of a normalized difference vegetative index (NDVI) as shown in Eq. 1.

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \quad (1)$$

The NDVI has been used extensively for the past 20 years as one of the standard VIs for application to crop canopies. It is important to understand the history of VIs and their limitations that were first proposed by the developers of the index. As an example of these limitations is the application of NDVI to various canopy parameters; however, the seasonal trajectory of this index shows a plateau during the vast portion of the growing season because the index becomes saturated at values between 0.9 and 1.0 and when the leaf area index of the canopy reaches values between 3.5 and 4.0 (Fig. 3). When compared to the NIR/RED ratio, there are differences throughout the growing season. Hatfield et al. (1984) evaluated the relationships between LAI, intercepted photosynthetically active radiation (IPAR), and VIs for durum wheat and found that NDVI and intercepted PAR were strongly related. They found that there were two relationships, one for the stage of accumulating leaf area and one for senescing leaf area. This was the first attempt to bring together the NDVI and intercepted light to explain the nonlinear nature of the NDVI and light interception. The comparison of NDVI across multiple data sets by different investigators was summarized by Hatfield et al. (2004). One of the utilities of this approach is being able to estimate intercepted PAR by crop canopies for use in crop growth models for biomass and yield estimates.

Compensating for the soil background in observations of crop canopies has posed a problem for many of the VIs and several modifications to the VIs have been developed to compensate for the soil background. There have been three forms of VIs developed to address the soil background problem. One of the first was proposed by Kauth and Thomas (1976) and used a linear combination of four wavebands using principal component analyses to estimate brightness, greenness, yellowness, etc. “Brightness” represents the magnitude of the reflected energy and could be considered a soil background line. “Greenness” represents an orthogonal plane to the soil line that contained information about vegetation. “Yellowness” is an additional plane to the soil and vegetation lines with particular sensitivity to senescent vegetation. These three parameters provide information content about plant crop canopies during the course of a growing season.

Another method of addressing the soil background was developed in the form of the perpendicular vegetative index (PVI) proposed by Richardson and Weigand (1977) using a statistical relationship given in Eq. 2.

$$PVI = (NIR - aRED - b) / (1 + a^2)^{1/2} \quad (2)$$

To simplify this relationship, Jackson (1983) proposed that PVI could be expressed as:

$$PVI = 0.647NIR - 0.763RED - 0.02 \quad (3)$$

As a further refinement, Huete (1988) developed ratio-based VIs to account for soil background through an approach he termed “soil-adjusted vegetative index” (SAVI) described as follows:

$$SAVI = \frac{NIR - RED}{NIR + RED + L} (1 + L) \quad (4)$$

The L term is an adjustment factor with a value of nearly 0.5 and was added to increase the sensitivity in the separation of canopy and soil at low ground cover amounts. The SAVI index was redefined by Rondeauz et al. (1996) who proposed an optimized SAVI (OSAVI) with the L term set equal to 0.16. Baret et al. (1989) provided an adjustment for soil background to the NDVI as the transformed soil-adjusted vegetative index (TSAVI) described as follows:

$$TSAVI = \frac{a(NIR - aRED - b)}{(RED + aNIR - ab)} \quad (5)$$

where a and b represent empirical coefficients obtained by regression fits to observed data. The changes to the VIs to compensate for soil background demonstrate the path that many VIs have taken toward improving the ability to detect differences in canopy responses.

Vegetative indices provide different representations of canopy parameters and there has been an evolution in these indices with the change from broadband to narrow

wave band or hyperspectral sensors. In many cases there has been a refinement in the wavelengths used in the VI but not a change in the functional description of the index. A summary of the different VIs and their application is presented by Hatfield et al. (2008).

Application to canopy parameters

Leaf chlorophyll content

The greenness of leaves as indicated by their chlorophyll content has emerged as one of the critical parameters in assessing the status of crop canopies. Leaves contain chlorophyll, Chl a and Chl b , essential pigments for the conversion of light energy to chemical energy. Solar radiation absorbed by leaves is a function of the photosynthetic pigment content and directly related to photosynthetic potential and primary production. Changes in leaf chlorophyll content are related to plant stress and senescence. One of the emerging applications of leaf chlorophyll observations is the estimation of the nutrient status because leaf nitrogen is incorporated into chlorophyll. There have been handheld chlorophyll absorbance meters developed to measure leaf transmittance at two wavelengths in the red (0.660 μ m) and near-infrared (NIR; 0.940 μ m) with the theoretical principles of these meters described by Markwell et al. (1995). The principles of the handheld meters have been extended to remote sensing systems that utilize different combinations of wavebands. An example of a canopy-level chlorophyll index is the modified chlorophyll absorption ratio index (MCARI) proposed by Daughtry et al. (2000) as:

$$MCARI = [(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})] \times (R_{700}/R_{670}) \quad (6)$$

This index was developed using three different wavebands that provide information about the chlorophyll absorption at 0.67 μ m relative to the green reflectance peak at 0.55 μ m and the near-infrared reflectance at 0.7 μ m. Using these three wavebands shows the maximum difference in reflectance for a leaf (Fig. 1). A refinement of the MCARI was proposed by Haboudane et al. (2002) as the transformed chlorophyll absorption ratio index (TCARI) that changed the form of the relationship as shown in Eq. 7:

$$TCARI = 3[(R_{700} - R_{670}) - 0.2(R_{700} - R_{550})] \times (R_{700}/R_{670}) \quad (7)$$

Gitelson et al. (2003) showed the possibility of methods to estimate chlorophyll content for plants based on simple ratio of NIR/RED reflectance $- 1$ or NIR/RED_{edge} reflectances $- 1$ and demonstrated the potential of these for canopy level assessments. The MCARI method was found to be sensitive to leaf area index, chlorophyll, leaf area-chlorophyll interactions, and background reflectance (Daughtry et al., 2000) and was especially sensitive to background reflectances when the canopy leaf area was small. Wu et al. (2008) provided a comparison of different

VIs for their ability to detect chlorophyll contents in winter wheat canopies and found the integrated methods that linked VIs sensitive to soil background (OSAVI) with chlorophyll indices (MCARI or TCARI) were the most robust. The estimation of leaf or canopy-level chlorophyll content provides valuable information about the canopy and potential response to nutrient stress or other stress that affects either Chl *a* or Chl *b* formation within the leaf. Zhang et al. (2008) applied these methods to predict the canopy chlorophyll content per unit ground area using hyperspectral sensors. Feret et al. (2008) combined a leaf optical model with the leaf chlorophyll information and leaf reflectance and transmittance to improve the estimation of chlorophyll content for canopies using remote sensing methods. These types of methods provide information about the canopy and are gaining wider use as a diagnostic tool for remote sensing of crop canopies.

Crop biomass

Crop biomass represents the total aboveground accumulation of plant material and is a tangible form of the net primary productivity of crop canopies. Remote sensing has been used to estimate dry matter accumulation or biomass estimation through a combination of NIR and red wavebands. These empirical fits have plant-specific relationships because of the difference in NIR reflectance among species, and this approach requires calibration for each crop and soil combination. There is a stronger relationship to green biomass with the NIR/RED combinations than to total biomass which includes stems, branches, or other nongreen matter.

Other approaches to estimating crop biomass have been to use a conversion factor of intercepted solar radiation to crop biomass using the following form of the relationship:

$$\text{Biomass} = \int \text{PAR} \bullet f\text{IPAR} \bullet \text{RUE} \bullet \Delta t \quad (8)$$

where PAR is the incident radiation, $f\text{IPAR}$ the fraction of intercepted photosynthetically active radiation (PAR) by the canopy, RUE the conversion efficiency of PAR to dry biomass, and Δt the time interval. The estimation of the intercepted values of PAR has taken on many different forms for this approach.

One component of biomass accumulation is the gross primary productivity (GPP) and Gitelson et al. (2006) found that GPP relates closely to total chlorophyll content in maize and soybean. The relationship algorithm for GPP estimation provided accurate estimates of midday GPP in both crops under rainfed and irrigated conditions. This approach has been evaluated using Landsat data to estimate canopy chlorophyll data and photosynthetically active radiation in a combination as proxies for GPP by Gitelson et al. (2012). They found these algorithms were able to estimate GPP accurately at three sites across the Midwest.

Intercepted solar radiation

Estimation of crop biomass is often based on intercepted light by crop canopies and is a critical component in plant growth models. Estimation of light interception by canopies from remotely sensed data would greatly aid in comparing management systems. Intercepted light by a crop canopy can be related to the accumulation of biomass and harvestable grain yield. Daily estimates of intercepted light can be obtained from extrapolation of observations of NDVI combined with daily total PAR obtained from a nearby meteorological station. It is possible to directly compare the performance of different cropping systems using this type of approach with confidence in the results. Observations collected over a wide range of crops and growing conditions suggest that LAI is best obtained from NIR/RED ratios while IPAR is best determined from NDVI.

One of the applications of intercepted light is to determine the rate of senescence of crop canopies. Examination of the rate of change in the NDVI shown in Fig. 3 can be utilized as a tool to examine how quickly plants are losing their physiological functions at the end of the growing season. This approach offers potential to determine if the rates of changes are different than expected and may indicate if there are some factors causing premature loss of green leaves in the canopy. This type of approach is often used for visual determination of premature changes in the canopy and could be easily determined from remote sensing platforms.

Crop ground cover

One of the components that is often evaluated for agricultural applications is the amount of ground covered by the crop canopy and is expressed as the fraction of ground area covered by the projection of standing leaf and stem area onto the ground surface. Changes in ground cover are often indicative of the health of the crop. Determination of ground cover provides a linkage between the growth of the crop and water use patterns of the crop since many evapotranspiration (ET) models use crop cover to relate potential ET to actual ET.

Maas (1998) proposed a method of estimating canopy ground cover in cotton that combined the overall reflectance of the scene and the individual reflectance values from the soil and the crop. He developed the following model for ground cover:

$$GC = (R_{\text{scene}} - R_{\text{soil}}) / (R_{\text{canopy}} - R_{\text{soil}}) \quad (9)$$

where GC is the fraction of ground cover, R_{scene} is the scene reflectance, R_{soil} is the soil reflectance, and R_{canopy} is the canopy reflectance. Scene reflectance is expressed as:

$$R_{\text{scene}} = R_{\text{canopy}}GC + R_{\text{soil}}(1 - GC) \quad (10)$$

He used reflectance values from either red (0.6–0.7 μm) or NIR (0.8–0.9 μm) for these relationships and

found that either waveband could be used. This method of estimating ground cover was independent of location and year. This method was not dependent upon empirical fits of the VI with plant parameters. Estimation of ground cover via remote sensing has proven to be fairly simple and not subject to problems associated with LAI or IPAR. The error in estimates of ground cover using these approaches has been on the order of $\pm 5\%$. This level of error is acceptable for agricultural applications that require ground cover estimates.

Canopy leaf area index

Estimation of the amount of green leaf area of a crop has been one of the primary approaches to understanding crop growth and development and quantifying the effects of different agronomic practices on crop growth. Leaf area index has been used to assess the ability of a plant to intercept light and LAI is used as one of the critical parameters in crop growth models. The NIR/RED ratio was found to be highly correlated with green LAI. However, there was a different form of the relationship of $LAI = a + b \text{ NIR/RED}$ for the growth and senescence portions of the growth cycle. Leaf area estimates for different crops across locations have been summarized by Wiegand and Hatfield (1988) and Wiegand et al. (1990). LAI for wheat was best estimated by the NIR/RED ratio or the TSAVI index using the linear relationships with coefficients of $LAI = a + b \text{ NIR/RED}$ and for TSAVI of $LAI = a + b \text{ TSAVI}$. Estimates of LAI are possible with simple linear models using NIR/RED reflectance. However, there is improvement when the same parameters are used but in a different form, with greatly improved sensitivity. These relationships are valid across a number of crops and agronomic practices within and among locations. To be useful for standard agronomic practices, these relationships need to be calibrated for a specific crop. The multisite comparisons for corn, wheat, and grain sorghum provide a degree of confidence that remote sensing measures can be adequately used to estimate LAI.

There is a relationship between LAI and light interception in plant canopies. One approach that has been proposed for estimating LAI is based on the relationships between fractional cover, f_C , and LAI using a relatively simple exponential relationship (Choudhury, 1987):

$$f_C = 1 - \exp(-\beta LAI) \quad (11)$$

where β is a function of the leaf angle distribution. He estimated β as 0.67 from an average of 18 crops. This method, although robust, has not been applied as often as NIR/RED ratios because the first step in this method is to obtain an estimate of ground cover or fractional cover and then incorporate it into Eq. 11. The multiple steps for this approach have contributed to the more widespread use of the simple regression models between VIs and LAI.

Crop yield

The estimation of green leaf area, intercepted light, or crop biomass is a foundation for the eventual estimation of harvestable crop yield. Yield estimation from remotely sensed data is generally accomplished with either reflectance-based or thermal-based models. These models can be characterized as some form of a relationship between rate of change in the VI and crop growth rate. Considerable variation in the response of the different models across locations has been observed because many of the approaches used empirical fits between yield and remotely sensed data. Deviations from normal grain-filling conditions either by soil, weather, or agronomic practices may not be detectable in simple VI relationships. Grain yields can be related to the rate of crop senescence. Rate of crop senescence can be determined from NDVI or NIR/RED ratios. Crop yield can be estimated from various remote sensing methods and generally involves some method of using the biomass or LAI values obtained from VIs as inputs into either crop growth models to predict yield or direct estimates of yield. Shanahan et al. (2001) used TSAVI, NDVI, and a Green NDVI (GNDVI) where the red reflectance replaced by green reflectance showed an improvement in yield estimation. They found GNDVI provided a slight improvement in yield estimates for the four corn hybrids grown with varying nitrogen (N) and water levels and the utility of this method is the spatial distribution of yield within fields. There continues to be studies reported that relate green, red, NIR, and NDVI values to examine field scale variation in grain yield. In these studies, yield is positively related to NIR and NDVI values and negatively related to green and red values. However, when a comparison is made across the studies there is a large variation in the coefficients in the statistical relationships between VI and grain yield.

There is a need for continued refinement of remote sensing approaches to estimate yield because of the importance in predicting large-scale crop yields or developing field scale maps of yield variation as a part of precision agriculture. One of the areas of potential improvement in yield predictions from remotely sensed observations will occur if we begin to critically evaluate the reasons for the lack of fit between observed and predicted data. Much of the variation can be explained by failure to have field validation data collected at the proper time or incorporation of the improper VIs into these relationships.

Utilization of thermal radiation

Thermal radiation emitted from crop or soil surfaces is representative of the temperature of the surface. However, the utilization of thermally infrared-derived temperatures for assessment of agricultural systems requires a comparison to air temperature. Crop stress methods using surface temperature are based on the comparison with air temperature because of the linkage to the sensible heat exchange of the crop surface. Leaves or canopies that

are cooler than the air temperature have a rapid evaporation rate while those that are warmer than air have a reduced evaporation rate. Plants that have adequate soil water supplies will tend to have leaves cooler than air and as the water supply becomes more limited the leaves will gradually become warmer than the surrounding air. This is premise of all approaches that utilize canopy temperatures that assess crop stress. Unlike the VIs, the use of thermal infrared information has been formed from indices derived from combinations of air and canopy temperatures as empirical relationships to crop yield or biomass based on a comparison to the expected values for a crop in a non-stressed environment. One of the uses of thermal infrared has been the estimation of evapotranspiration from crop canopies.

Variation of the simple canopy: Air temperature difference ($T_c - T_a$) across crops and climates invoked a series of studies that led Idso et al. (1981) to derive an empirical model for canopy stress, while Jackson et al. (1981) derived the more theoretical relationship based on the energy balance of a canopy. The empirical model for CWSI is described as:

$$\text{CWSI} = \frac{dT - \text{MIN}}{\text{MAX} - \text{MIN}} \quad (12)$$

where dT is $T_c - T_a$, MIN is the non-stressed baseline given as $a + b$ (Vapor Pressure Deficit, VPD), and MAX the upper limit of $T_c - T_a$ when the canopy is no longer transpiring. Values for MIN are obtained by measuring T_c throughout a day to obtain the data necessary for the regression equation. Values for MAX are constant across a range of VPDs and are observed in studies in which canopies are not completely water-stressed. The more critical relationship to define in Eq. 12 is the non-stressed lower baseline. Jackson et al. (1981) derived these curves from the relationship as follows:

$$\text{CWSI} = 1 - \frac{E}{E_p} = \frac{\gamma \left(1 + \frac{r_c}{r_a}\right) - \gamma \left(1 + \frac{r_{cp}}{r_a}\right)}{\Delta + \gamma \left(1 + \frac{r_c}{r_a}\right)} \quad (13)$$

where the quantity r_c/r_a is expressed as:

$$\frac{r_c}{r_a} = \frac{\frac{r_a R_n}{p c_p} - (T_c - T_a)(\Delta + \gamma) - (e_a^* - e_a)}{\gamma \left[(T_c - T_a) - \frac{r_a R_n}{p c_p} \right]} \quad (14)$$

and E is actual evaporation, E_p potential evaporation, r_{cp} the canopy resistance of a well-watered canopy, r_c the actual canopy resistance, r_a the aerodynamic resistance to sensible heat transfer, R_n the net radiation, Δ the slope of saturation curve, γ the psychrometric constant, e_a^* the saturation vapor pressure, and e_a the actual vapor pressure of the air. There are differences between Eqs. 13 and 14; however, one can derive the theoretical shape of Eq. 12 from the assumptions of a well-watered and a completely stressed or non-transpiring canopy.

Both of these approaches demonstrated the universal application of these relationships. The crop water stress index (CWSI) has become one of the most widely used methods for quantifying crop stress. The efforts to use these approaches have been summarized in Hatfield et al. (2008).

One of the issues with use of canopy temperatures that is similar to using VIs from reflected radiation is the problem of incomplete canopy cover. In the application of canopy temperatures, the problem of incomplete ground cover is more critical because of the potential differences in the temperatures between the soil and the crop. The use of the average temperature based on a fractional ground cover may not provide the correct average compared to the fractional representation of reflectance values. The development of multiscale approaches that combine thermal, visible, and near-infrared imagery from multiple satellites to partition the fluxes between the soil and canopy offer the potential for future improvements in the use of surface temperature at scales ranging from 1 m to 10 km (Anderson et al., 2007). This type of method shows further refinement in the ability to use remote sensing as an assessment tool for ground-based observations as well as a method for regional-scale measurements.

Application to agricultural problems

Crop nutrient status

Estimation of the crop nutrient status is one of the goals of remote sensing and critical to agricultural production because of the need to optimize crop production relative to the nutrient inputs and reduce the potential environmental impacts from agricultural production systems. The detection of chlorophyll content that was discussed earlier has formed the basis for the development of tools to detect either nitrogen or phosphorus stress in crop canopies. There have been studies that have shown that improvements in nutrient management are possible with remote sensing tools that are calibrated to detection of nutrient stress. These efforts will continue to develop and provide valuable information for management decisions.

Pests and other stresses

There are a variety of potential stresses from a variety of other biotic factors. Detection of biotic stresses using remote sensing is not as advanced as the estimation of canopy parameters. There are a few examples of how remote sensing can detect pests. Riedell and Blackmer (1999) found that aphids (*Diuraphis noxia*, Mordvilko) and greenbugs (*Schizaphis graminum*, Rondani) caused changes in the 625–635 μm and 680–695 μm wavebands because of their effect on leaf chlorophyll status. For this approach to effectively detect stress, a comparison would have to be made to a non-stressed plant. One of the merits of this approach is the ability to detect the

presence of aphids in parts of the field because these insects generally do not uniformly affect a field but occur in isolated areas.

Many leaf diseases cause changes in leaf reflectance and many of these observations have been made for a number of years. One of the early reports was by Blazquez and Edwards (1983) who showed it was possible to detect disease occurrence in both tomato and potato leaves using spectral reflectance. Comparisons that relate disease occurrence, timing, and location information are likely to provide some method of discriminating between disease and healthy leaves at the canopy level. Detection of insects and diseases using a combination of spectral signatures has been conducted at the leaf and canopy scale but the routine extension to the field scale has not been demonstrated.

Weeds within canopies provide another source of confusion with remotely sensed data. These effects are unlike insects or diseases because spectral signatures of weed leaves are similar to crop canopies. Medlin et al., (2000) used green (0.535–0.545 μm), red (0.69–0.7 and 0.715–0.725 μm), and NIR (0.835–0.845 μm) wavebands to detect *Ipomoea lacunosa* L., *Senna obtusifolia* (L.) Irwin et Barnaby, and *Solanum carolinense* L. in soybean canopies. They could detect these infestations within the soybean canopy with 75 % accuracy; however, the greatest difficulty was in the detection of weed-free areas in the field. Their ability to detect weeds was due to differences in reflectance for these wavebands between weeds and the crop. Detection of weeds within various crop and rangeland or pasture canopies has been reported for *Ericameria austrotexana* M. Johnston. (Anderson et al., 1993), *Euphorbia esula* L. (Everitt et al., 1995), and *Chrysanthemum leucanthemum* L. (Lass and Callihan, 1997). Other studies have found only limited success in differentiating crops and weeds because of the similarity of spectral reflectance of weeds and crops (Richardson et al., 1985; Menges et al., 1985). They found they were able to detect plant community areas later in the season due to increased ground cover of weed-crop mixture.

Observations for current research suggest two potential avenues for detection of weeds within crop canopies. First is the direct detection of the weed within the crop by separating the spectral signatures for weeds versus crops. Second is the change in field variability patterns because weeds are not evenly distributed across fields and often aggregate in patches across fields similar to detection of insects or diseases within fields. Both of these approaches have not been extensively explored or evaluated with remote sensing tools. The improvements in both the spatial resolution of reflectance data and the temporal frequency provide opportunities for integrated studies of the remote sensing community with agronomists, weed scientists, entomologists, nematologists, and plant pathologists that will provide new techniques for detection of stresses in crop communities.

Conclusions

Remote sensing for agriculture has had a rich history of providing information about crop canopies. The utilization of reflectance data is enhanced through the incorporation into VIs, and the continual refinement of these VIs for application beyond the estimation of canopy parameters will continue to improve the knowledge base of crop changes throughout the year. The utility of remote sensing to move beyond observation to potential field-scale management by providing more detailed information on the spatial patterns within fields and the value of that information will increase the richness of the data and increase the value for the producer. Increasing the application of these relationships for all crops beyond the agronomic systems and into vegetables, vines, and fruits will require improved understanding of the relationships between reflectance and emittance and canopy parameters. This field continues to evolve and offers the potential for many new areas of investigation.

Bibliography

- Allen, W. A., Gausman, H. W., Richardson, A. J., and Thomas, J. R., 1969. Interaction of isotropic light with a compact plant leaf. *Journal of the Optical Society of America*, **59**, 1376–1379.
- Anderson, G. L., Everitt, J. H., Richardson, A. J., and Escobar, D. E., 1993. Using satellite data to map false broomweed (*Ericameria austrotexana*) infestations on south Texas rangelands. *Weed Technology*, **7**, 865–871.
- Anderson, M. C., Kustas, W. P., and Norman, J. M., 2007. Upscaling flux observations from local to continental scales using thermal remote sensing. *Agronomy Journal*, **99**, 240–254.
- Baret, F., Guyot, G., and Major, D. J., 1989. TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation. In *Proceedings of the IGARRS '0/12th Canadian Symposium on Remote Sensing*, Vancouver, Canada, Vol. 3, pp. 1355–1358.
- Barnes, E. M., Sudduth, K. A., Hummel, J. W., Leach, S. M., Corwin, D. L., Yeng, C., Daughtry, C. S. T., and Bausch, W. C., 2003. Remote- and ground-based sensor techniques to map soil properties. *Photogrammetric Engineering & Remote Sensing*, **69**, 619–630.
- Blazquez, C. H., and Edwards, G. J., 1983. Infrared color photography and spectral reflectance of tomato and potato diseases. *Journal of Applied Photographic Engineering*, **9**, 33–37.
- Choudhury, B. J., 1987. Relationships between vegetation indices, radiation absorption, and net photosynthesis evaluated by a sensitivity analysis. *Remote Sensing of Environment*, **22**, 209–233.
- Daughtry, C. S. T., Walthall, C. L., Kim, M. S., Brown de Colstoun, E., and McMurtrey, J. E., III, 2000. Estimating corn leaf chlorophyll content from leaf and canopy reflectance. *Remote Sensing of Environment*, **74**, 229–239.
- Deering, D. W., 1978. *Rangeland Reflectance Characteristics Measured by Aircraft and Spacecraft Sensors*. PhD dissertation, College Station, TX, Texas A&M University, 338 pp.
- Doraiswamy, P. C., Moulin, S., and Cook, P. W., 2003. Crop yield assessment from remote sensing. *Photogrammetric Engineering & Remote Sensing*, **69**, 665–674.
- Everitt, J. H., Anderson, G. L., Esobar, D. E., Davis, M. R., Spencer, N. R., and Andrascik, R. J., 1995. Use of remote sensing for detecting and mapping leafy spurge (*Euphorbia esula*). *Weed Technology*, **9**, 599–609.

- Feret, J. B., Francois, C., Asner, G. A., Gitelson, A. A., Martin, R. E., Bidet, L. P. R., Ustin, S. L., le Maire, G., and Jacquemoud, S., 2008. PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sensing of Environment*, **112**, 3030–3043.
- Gates, D. M., Keegan, H. J., Schleter, J. C., and Weidner, V. R., 1965. Spectral properties of plants. *Applied Optics*, **4**, 11–20.
- Gausman, H. W., and Allen, W. A., 1973. Optical parameters of leaves of 30 plant species. *Plant Physiology*, **52**(1), 57–62.
- Gausman, H. W., Allen, W. A., Myers, V. I., and Cardenas, R., 1969. Reflectance and internal structure of cotton leaves *Gossypium hirsutum* L. *Agronomy Journal*, **61**, 374–376.
- Gitelson, A. A., Gritz, U., and Merzlyak, M. N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, **160**(3), 271–282.
- Gitelson, A. A., Viña, A., Verma, S. B., Rundquist, D. C., Arkebauer, T. J., Keydan, G., Leavitt, B., Ciganda, V., Burba, G. G., and Suyker, A. E., 2006. Relationship between gross primary production and chlorophyll content in crops: implications for the synoptic monitoring of vegetation productivity. *Journal of Geophysical Research*, **111**, D08S11, doi:10.1029/2005JD006017.
- Gitelson, A. A., Peng, Y., Masek, J. G., Rundquist, D. C., Verma, S., Suyker, A., Baker, J. M., Hatfield, J. L., and Meyers, T., 2012. Remote estimation of crop gross primary productivity with Landsat data. *Remote Sensing of Environment*, **121**, 404–412.
- Haboudane, D., John, R., Millera, J. R., Tremblay, N., Zarco-Tejada, P. J., and Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, **81**, 416–426.
- Hatfield, J. L., Asrar, G., and Kanemasu, E. T., 1984. Intercepted photosynthetically active radiation estimated by spectral reflectance. *Remote Sensing of Environment*, **14**, 65–75.
- Hatfield, J. L., Prueger, J. H., and Kustas, W. P., 2004. Remote sensing of dryland crops. In Ustin, S. L. (ed.), *Remote sensing for natural resource management and environmental monitoring*, 3rd edn. Hoboken, NJ: Wiley. Manual of remote sensing, Vol. 4, Chap. 10, pp. 531–568.
- Hatfield, J. L., Gitelson, A. A., Schepers, J. S., and Walthall, C. L., 2008. Application of spectral remote sensing for agronomic decisions. *Agronomy Journal*, **100**, S-117–S-131.
- Huete, A. R., 1988. A soil-adjusted vegetative index (SAVI). *Remote Sensing of Environment*, **25**, 295–309.
- Hunt, E. R., Everitt, J. H., Ritchie, J. C., Moran, M. S., Booth, D. T., Anderson, G. L., Clairk, P. E., and Seyfried, M. S., 2003. Applications and research using remote sensing for rangeland management. *Photogrammetric Engineering & Remote Sensing*, **69**, 675–693.
- Idso, S. B., Jackson, R. D., Pinter, P. J., Jr., Reginato, R. J., and Hatfield, J. L., 1981. Normalizing the stress-degree-day parameter for environmental variability. *Agricultural Meteorology*, **24**, 45–55.
- Jackson, R. D., 1983. Spectral indices in n-space. *Remote Sensing of Environment*, **13**, 409–421.
- Jackson, R. D., and Huete, A. R., 1991. Interpreting vegetation indices. *Preventive Veterinary Medicine*, **11**, 185–200.
- Jackson, R. D., Idso, S. B., Reginato, R. J., and Pinter, P. J., Jr., 1981. Canopy temperature as a crop water stress indicator. *Water Resources Research*, **17**, 1133–1138.
- Jordan, C. F., 1969. Derivation of leaf area index from quality of light on the forest floor. *Ecology*, **50**, 663–666.
- Kauth, R. J., and Thomas, G. S., 1976. The tasseled cap-A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*. West Lafayette, IN: Purdue University, pp. 41–51.
- Kustas, W. P., French, A. N., Hatfield, J. L., Jackson, T. J., Moran, M. S., Rango, A., Ritchie, J. C., and Schmugge, T. J., 2003. Remote sensing research in hydrometeorology. *Photogrammetric Engineering & Remote Sensing*, **69**, 631–646.
- Lass, L. W., and Callihan, R. H., 1997. The effect of phenological stage on detectability of yellow hawkweed (*Hieracium pratense*) and oxeye daisy (*Chrysanthemum leucanthemum*) with remote multispectral digital imagery. *Weed Technology*, **11**, 248–256.
- Maas, S. J., 1998. Estimating cotton canopy cover from remotely sensed scene reflectance. *Agronomy Journal*, **90**, 384–388.
- Markwell, J., Osterman, J. C., and Mitchell, J. L., 1995. Calibration of the Minolta SPAD-502 leaf chlorophyll meter. *Photosynthesis Research*, **46**, 467–472.
- Medlin, C. R., Shaw, D. R., Gerard, P. D., and LaMastus, F. E., 2000. Using remote sensing to detect weed infestations in *Glycine max*. *Weed Science*, **48**, 393–398.
- Menges, R. M., Nixon, P. R., and Richardson, A. J., 1985. Light reflectance and remote sensing of weeds in agronomic and horticultural crops. *Weed Science*, **33**, 569–581.
- Moran, S., Fitzgerald, G., Rango, A., Walthall, C., Barnes, E., Bausch, W., Clarke, T., Daughtry, C., Everitt, J., Escobar, D., Hatfield, J., Havstad, K., Jackson, T., Kitchen, N., Kustas, W., McGuire, M., Pinter, P., Jr., Sudduth, K., Schepers, J., Schmugge, T., Starks, P., and Upchurch, D., 2003. Sensor development and radiometric correction for agricultural applications. *Photogrammetric Engineering & Remote Sensing*, **69**, 705–718.
- Pinter, P. J., Hatfield, J. L., Schepers, J. S., Barnes, E. M., Moran, M. S., Daughtry, C. S., and Upchurch, D. R., 2003. Remote sensing for crop management. *Photogrammetric Engineering and Remote Sensing*, **69**, 647–664.
- Richardson, A. J., and Weigand, C. L., 1977. Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, **43**, 1541–1552.
- Richardson, A. J., Menges, R. M., and Nixon, P. R., 1985. Distinguishing weed from crop plants using video remote sensing. *Photogrammetric Engineering and Remote Sensing*, **51**, 1785–1790.
- Riedell, W. E., and Blackmer, T. M., 1999. Leaf reflectance spectra of cereal aphid-damaged wheat. *Crop Science*, **39**, 1835–1840.
- Ritchie, J. C., Zimba, P. V., and Everitt, J. H., 2003. Remote sensing techniques to assess water quality. *Photogrammetric Engineering & Remote Sensing*, **69**, 695–704.
- Rondeauz, G., Steven, M., and Baret, F., 1996. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, **55**, 95–107.
- Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., Schlemmer, M. R., and Major, D. J., 2001. Use of remote-sensing imagery to estimate corn grain yield. *Agronomy Journal*, **93**, 583–589.
- Tanner, C. B., 1963. Plant temperature. *Agronomy Journal*, **55**, 210–211.
- Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**, 127–150.
- Wiegand, C. L., and Hatfield, J. L., 1988. The spectral-agronomic multisite-multicrop analyses (SAMMA) project. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **27**(B7), 696–706.
- Wiegand, C. L., Gerbermann, A. H., Gallo, K. P., Blad, B. L., and Dusek, D., 1990. Multisite analyses of spectral-biophysical data for corn. *Remote Sensing of Environment*, **33**, 1–16.
- Woolley, J. T., 1971. Reflectance and transmittance of light by leaves. *Plant Physiology*, **47**, 656–662.
- Wu, C., Nu, Z., Tang, Q., and Huang, W., 2008. Estimating chlorophyll content from hyperspectral vegetative indices:

modeling and validation. *Agricultural and Forest Meteorology*, **148**, 1230–1241.

Zhang, V., Chen, J. M., Miller, J. R., and Noland, T. L., 2008. Leaf chlorophyll content retrieval from airborne hyperspectral remote sensing imagery. *Remote Sensing of Environment*, **112**, 3234–3247.

Cross-references

[Crop Stress](#)

[Vegetation Indices](#)

AIR POLLUTION

Annmari Elderling

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

Air quality

Definitions

Ozone. A gaseous form of oxygen with three atoms per molecule. Ozone is a bluish gas that is harmful to breathe. Nearly 90 % of the Earth's ozone is in the stratosphere and is referred to as the ozone layer. High concentrations of ozone in the lower troposphere make up one of the components of air pollution.

Troposphere. The portion of the atmosphere, which extends outward about 10–20 km from the Earth's surface and in which generally temperature decreases rapidly with altitude, clouds form, and convection and weather are active.

Stratosphere. The region of the Earth's atmosphere extending from the tropopause to about 50 km (31 miles) above the Earth's surface. The stratosphere is characterized by the presence of ozone gas (in the ozone layer) and by temperatures that rise slightly with altitude, due to the absorption of ultraviolet radiation by ozone.

Aerosols. An aerosol is a suspension of fine solid particles or liquid droplets in a gas. Some examples are pollution haze, smog, and oceanic haze.

Introduction

Air pollution refers to increased concentrations of gases or particles that are harmful to human health, vegetation, and welfare. Often, this phrase refers to gases that are regulated as air pollutants by agencies such as the Environmental Protection Agency in the United States. Specifically, the regulated pollutants are ozone, carbon monoxide, sulfur dioxide, nitrogen oxides, lead, and particulate matter. Lead is not measured by remote sensing and will not be discussed here. The study of aerosols with remote sensing is discussed in its own entry under Aerosols.

The Earth's atmosphere is described in layers, starting with the troposphere, which extends from the surface to 10–15 km, depending on location. Since humans and our biosphere occupy only the lowest layers of the atmosphere, much of the focus of studying air pollutants remains on the lowest, near-surface, layers of the troposphere. Still, there is reason to measure all layers of the troposphere, as transport of pollutants and chemical processing occurs throughout the troposphere and can impact the concentrations in the lowest layers. The stratosphere is the layer above the troposphere, and it extends from the tropopause to about 50 km above the surface. Trace gases and pollutants are distributed unequally in Earth's atmosphere, with some pollutants, like CO and SO₂, present at their highest concentrations in the troposphere. Other gases, such as ozone and NO₂, are present in both the stratosphere and the troposphere, so measuring the portion relevant to air pollution requires techniques that differentiate the troposphere and stratosphere. Remote sensing provides the capability to make measurements over a large fraction of the globe but with some limits on our ability to see the chemicals of interest in the troposphere.

This field has matured over the last 3 decades, and there are some excellent overview papers in the scientific literature that provide a history and review (Fishman et al., 2008; Martin, 2008).

This entry reviews total column measurements of CO and ozone and then discusses approaches to measure tropospheric ozone and CO. Finally, other tropospheric pollutants, namely, SO₂ and NO₂, are discussed. Note that this entry was written in 2010 and reflects the measurements up to that point in time.

Measurement of total columns of pollutants

The Measurement of Air Pollution from Satellites (MAPS) experiment was flown on the Space Shuttle four times between 1981 and 1994, and using infrared channels, it provided the first global views of CO concentrations. The technique of the Measurement of Air Pollution from Satellites (MAPS) measurement (correlation radiometer in the infrared) is most sensitive to enhancements in the mid- to upper troposphere. Measurements taken over 8–10 day periods clearly showed the high CO concentrations that came from biomass burning activities in October of 1984 and 1994, and the April 1994 measurements showed that there was some enhancement of CO across the whole Northern Hemisphere (Connors et al., 1999; Reichle et al., 1999).

A number of sensors have been used to make global measurements of CO with similar wavelengths and similar vertical sensitivity. The Japanese infrared instrument, Interferometric Monitoring of Greenhouse Gases (IMG), made global CO measurements for several months in 1996–1997 (Hadji-Lazaro et al., 1999). The Measurements of Pollution in the Troposphere (MOPITT) instrument on NASA's EOS Terra was launched in 1999 and

now has over 10 years of CO measurement data (Deeter et al., 2003). The Atmospheric Infrared Sounder (AIRS) instrument on EOS Aqua, Tropospheric Emission Spectrometer (TES) on EOS Aura, and Infrared Atmospheric Sounding Interferometer (IASI) on CNES/EUMETSAT's METOP platform have all demonstrated CO measurements from infrared channels (McMillan et al., 2005; Rinsland et al., 2006; Clerbaux et al., 2009).

In addition to CO measurements, the total column of ozone has been measured from space for decades. The Total Ozone Mapping Spectrometer (TOMS) instrument was flown in 1978, with a focus on measuring the total column ozone and contributing to our understanding of the polar ozone holes. A number of techniques have been developed to make an estimate of the amount of ozone in the stratosphere (generally from limb sounders) and take the difference of the total column ozone and the stratospheric amount to get an estimate of the tropospheric ozone. This is generally referred to as the tropospheric ozone residual (see Fishman et al., 1990) and was originally applied to TOMS total columns and SAGE stratospheric profiles. This technique has been adapted to use other column measurements (such as the Ozone Monitoring Instrument (OMI)) and other stratospheric estimates (such as MLS). While this technique provided some of the first global views of the tropospheric ozone distribution, it has limited sensitivity to the ozone in the lower troposphere.

Measurements of tropospheric ozone

An innovation in the measurement of tropospheric pollutants, especially ozone, came about with the EOS Aura Tropospheric Emission Spectrometer (TES). The TES instrument measures with very high spectral resolution (0.1 cm^{-1}) in the infrared, which allows for differentiation of tropospheric and stratospheric ozone. In cases where the surface temperatures are warm and there is a large amount of ozone, TES can differentiate the amount of ozone in the upper and lower troposphere (Worden et al., 2007; Jourdain et al., 2007). TES has provided global views of tropospheric ozone, and with the simultaneous CO measurements, the data has been used to estimate the impact of large biomass burning events on tropospheric pollution (Logan et al., 2008) and of the long-range transport of anthropogenic pollution (Zhang et al., 2006).

Ozone retrievals from infrared measurements with coarser spectral resolution (AIRS and IASI) have shown very limited sensitivity to ozone below 300 hPa (Pittman et al., 2009). A number of research algorithms have been applied to the IASI data for ozone profile retrievals, with some differences in sensitivity and biases, but they consistently show limited sensitivity below 5 km (Keim et al., 2009).

A methodology to retrieve tropospheric columns of ozone directly from satellite measurements in the ultraviolet and visible has been reported (Chance et al., 1997). This technique was demonstrated with global retrieval of tropospheric ozone column from the Global Ozone

Mapping Experiment (GOME) satellite data in Liu et al., (2006). GOME (Burrows et al., 1999) was launched aboard the European Space Agency (ESA) European Remote Sensing satellite ERS-2 in 1995. This technique requires a good estimate of the tropopause height and can have limited sensitivity to lower tropospheric ozone. This technique has since been applied to the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) and GOME-2. SCIAMACHY (Bovensmann et al., 1999) was launched aboard ESA's Envisat in 2002, and GOME-2 (Callies et al., 2003) was launched aboard ESA's MetOp-A in 2006. Most recently, it has been applied to the NASA EOS OMI data (Liu et al., 2009).

Measurements of tropospheric CO

The MOPITT instrument was designed to exploit both thermal infrared (TIR) and near-infrared (NIR) absorption features of CO. In principal, the use of these two bands together should provide better discrimination of the tropospheric distribution of CO, providing more sensitivity to the surface than any other measurement of CO. Although MOPITT has been making measurement with both of these bands, it has been difficult to use the NIR measurements, but progress has been made on this front (Deeter et al., 2009). Recently (Worden et al., 2009), CO retrievals that combined the TIR and NIR measurements from MOPITT have been demonstrated, and they do show more sensitivity to the surface.

Other pollutants: NO₂ and SO₂

Nitrogen dioxide (NO₂) is present in the stratosphere and troposphere, but because of significant tropospheric sources and a lifetime in the troposphere that is a day or two, the variability of NO₂ is driven by tropospheric variability. This has allowed satellite measurements of NO₂ to be used to estimate the tropospheric column, by making an adjustment for the stratospheric amount, based on the column over a clean region (generally over the central area of oceans). As with ozone, this technique has been applied to many satellite measurements – GOME, SCIAMACHY, and OMI (Martin et al., 2002; Richter and Burrows, 2002; Beirle et al., 2003; Boersma et al., 2004; Bucsel et al., 2006).

Sulfur dioxide (SO₂) has been measured from a variety of UV remote sensing instruments, including TOMS (Krueger, 1983; Krueger et al., 1995), SBUV/2 (McPeters, 1993), GOME (Eisinger and Burrows, 1998), OMI (Carn et al., 2007; Yang et al., 2007; Krotkov et al., 2008), and SCIAMACHY (Lee et al., 2008). The first sulfur dioxide measurements were of volcanic eruptions, and that is the focus of most remote sensing measurements. Volcanic SO₂ is often injected into the upper troposphere and lower stratosphere in large eruptions, and this provides a strong signal in comparison to the background. Eisinger and Burrows, (1998) showed that they could measure tropospheric sulfur dioxide enhancements due to coal burning

power plants. Emissions from copper smelters and power plants have been measured with OMI (Krotkov et al., 2008), which has two orders of magnitude improvement in the minimum detectable SO₂ mass relative to TOMS, due to a smaller footprint and improvements in the wavelengths used in the retrieval.

Conclusions

A number of remote sensing techniques have been developed to detect and quantify key air pollutants in the troposphere. Using ultraviolet, visible, and infrared wavelengths, many species have now been measured over a number of years. More recently, TES on EOS Aura has demonstrated that with high spectral resolution in the infrared, stratospheric ozone can be differentiated from tropospheric ozone, and under polluted conditions, the upper and lower tropospheric ozone can be differentiated. Recent work on MOPITT data has shown that CO measurements with sensitivity near the surface can be extracted from using the 4.6 and 2.3 μm bands together.

Acknowledgment

The research to prepare this entry was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Bibliography

- Beirle, S., Platt, U., Wenig, M., and Wagner, T., 2003. Weekly cycle of NO₂ by GOME measurements: a signature of anthropogenic sources. *Atmospheric Chemistry and Physics*, **3**, 2225–2232.
- Boersma, K. F., Eskes, H. J., and Brinksma, E. J., 2004. Error analysis for tropospheric NO₂ retrieval from space. *Journal of Geophysical Research*, **109**, D043011, doi:10.1029/2003JD003962.
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H., 1999. SCIAMACHY: mission objectives and measurement modes. *Journal of Atmospheric Science*, **56**, 127–150.
- Bucsela, E. J., Celarier, E. A., Wenig, M. O., Gleason, J. F., Veefkind, J. P., Boersma, K. F., and Brinksma, E. J., 2006. Algorithm for NO₂ vertical column retrieval from the ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1245–1258.
- Burrows, J. P., Weber, M., Buchwitz, M., Razonov, V., Ladstätter, A., Richter, A., De Beek, R., Hoogen, R., Bramsdte, D., Eichmann, K. U., Eisenger, M., and Perner, D., 1999. The global ozone monitoring experiment (GOME): mission concept and first scientific results. *Journal of Atmospheric Sciences*, **56**, 151–175.
- Callies, J., Corpaccioli, E., Eisinger, M., Lefebvre, A., Munro, R., Perez-Albinan, A., Ricciarelli, B., Calamai, L., Gironi, G., Veratti, R., Otter, G., Eschen, M., and van Riel, L., 2003. GOME-2 the ozone instrument on-board the European METOP satellites. *Proceedings of SPIE*, **5158**, 1–11.
- Carn, S. A., Krueger, A. J., Krotkov, N. A., Yang, K., and Levelt, P. F., 2007. Sulfur dioxide emissions from Peruvian copper smelters detected by the ozone monitoring instrument. *Geophysical Research Letters*, **34**, L09801, doi:10.1029/2006GL029020.
- Chance, K. V., Burrows, J. P., Perner, D., and Schneider, W., 1997. Satellite measurements of atmospheric ozone profiles, including tropospheric ozone, from ultraviolet/visible measurements in the nadir geometry: a potential method to retrieve tropospheric ozone. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **51**, 461–476.
- Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F., 2009. Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder. *Atmospheric Chemistry and Physics*, **9**, 6041–6054.
- Connors, V. S., Gormsen, B. B., Nolf, S., and Reichle, H. G., Jr., 1999. Spaceborne observations of the global distribution of carbon monoxide in the middle troposphere during April and October 1994. *Journal of Geophysical Research*, **104**, 21455–21470.
- Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J.-F., Ho, S.-P., Yudin, V., Attié, J.-L., Packman, D., Chen, J., Mao, D., and Drummond, J. R., 2003. Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument. *Journal of Geophysical Research Atmospheres*, **108**, 4399.
- Deeter, M. N., Edwards, D. P., Gille, J. C., and Drummond, J. R., 2009. CO retrievals based on MOPITT near-infrared observations. *Journal of Geophysical Research*, **114**, D04303, doi:10.1029/2008JD010872.
- Eisinger, M., and Burrows, J. P., 1998. Tropospheric sulfur dioxide observed by the ERS-2 GOME instrument. *Geophysical Research Letters*, **25**, 4177–4180.
- Fishman, J., Watson, C. E., Larsen, J. C., and Logan, J. A., 1990. Distribution of tropospheric ozone determined from satellite data. *Journal of Geophysical Research*, **95**, 3599–3617.
- Fishman, J., et al., 2008. Remote sensing of tropospheric pollution from space. *Bulletin of the American Meteorological Society*, **89**, 805–821, doi:10.1175/2008BAMS2526.1.
- Hadji-Lazaro, J., Clerbaux, C., and Thiria, S., 1999. An inversion algorithm using neural networks to retrieve atmospheric CO total columns from high-resolution nadir radiances. *Journal of Geophysical Research*, **104**, 23841–23854.
- Jourdain, L., Worden, H. M., Worden, J. R., Bowman, K., Li, Q., Eldering, A., Kulawik, S. S., Osterman, G., Boersma, K. F., Fisher, B., Rinsland, C. P., Beer, R., and Gunson, M., 2007. Tropospheric vertical distribution of tropical Atlantic ozone observed by TES during the northern African biomass burning season. *Geophysical Research Letters*, **34**, L04810, doi:10.1029/2006GL028284.
- Keim, C., et al., 2009. Tropospheric ozone from IASI: comparison of different inversion algorithms and validation with ozone sondes in the northern middle latitudes. *Atmospheric Chemistry and Physics (Discussion)*, **9**, 11441–11479.
- Krotkov, N. A., et al., 2008. Validation of SO₂ retrievals from the ozone monitoring instrument over NE China. *Journal of Geophysical Research*, **113**, D16S40, doi:10.1029/2007JD008818.
- Krueger, A. J., 1983. Sighting of El Chich'on sulfur dioxide clouds with the Nimbus 7 total ozone mapping spectrometer. *Science*, **220**, 1377–1379.
- Krueger, A. J., Walter, L. S., Bhartia, P. K., Schnetzler, C. C., Krotkov, N. A., Sprod, I., and Bluth, G. J. S., 1995. Volcanic sulfur dioxide measurements from the total ozone mapping spectrometer instruments. *Journal of Geophysical Research*, **D100**, 14057–14076.
- Lee, C., Richter, A., Weber, M., and Burrows, J. P., 2008. SO₂ retrieval from SCIAMACHY using the weighting function DOAS (WFDOAS) technique: comparison with standard DOAS retrieval. *Atmospheric Chemistry and Physics*, **8**, 6137–6145.
- Liu, X., et al., 2006. First directly retrieved global distribution of tropospheric column ozone from GOME: comparison with the GEOS-CHEM model. *Journal of Geophysical Research*, **111**, D02308, doi:10.1029/2005JD006564.
- Liu, X., Bartia, P. K., Chance, K., Spurr, R. J. D., and Kurosu, T. P., 2010. Abstract ozone profile retrievals from the ozone

- monitoring instrument. *Atmospheric Chemistry and Physics Discussions*, **9**, 22693–22738.
- Logan, J. A., Megretskaia, I., Nassar, R., Murray, L. T., Zhang, L., Bowman, K. W., Worden, H. M., and Luo, M., 2008. Effects of the 2006 El Niño on tropospheric composition as revealed by data from the tropospheric emission spectrometer (TES). *Geophysical Research Letters*, **35**, L03816, doi:10.1029/2007GL031698.
- Martin, R. V., 2008. Satellite remote sensing of surface air quality. *Atmospheric Environment*, **42**, 7823–7843.
- Martin, R.V., and Coauthors, 2002. An improved retrieval of tropospheric nitrogen dioxide from GOME. *Journal of Geophysical Research*, **107**, 4437, doi:10.1029/2001JD001027
- McMillan, W. W., Barnett, C., Strow, L., Chahine, M. T., McCourt, M. L., Warner, J. X., Novelli, P. C., Korontzi, S., Maddy, E. S., and Datta, S., 2005. Daily global maps of carbon monoxide from NASA's atmospheric infrared sounder. *Geophysical Research Letters*, **32**, L11801, doi:10.1029/2004GL021821.
- McPeters, R. D., 1993. The atmospheric SO₂ budget for Pinatubo de-ri-ved from NOAA-11 SBUV/2 spectral data. *Geophysical Research Letters*, **20**, 1971–1974.
- Pittman, J. V., Pan, L. L., Wei, J. C., Irion, F. W., Liu, X., Maddy, E. S., Barnett, C. D., Chance, K., and Gao, R.-S., 2009. Evaluation of AIRS, IASI, and OMI ozone profile retrievals in the extratropical tropopause region using in situ aircraft measurements. *Journal of Geophysical Research*, **114**, D24109, doi:10.1029/2009JD012493.
- Reichle, H. G., et al., 1999. Space shuttle based global CO measurements during April and October 1994, MAPS instrument, data reduction, and data validation. *Journal of Geophysical Research*, **104**, 21443–21454.
- Richter, A., and Burrows, J. P., 2002. Tropospheric NO₂ from GOME measurements. *Advances in Space Research*, **29**, 1673–1683.
- Rinsland, C. P., Luo, M., Logan, J. A., Beer, R., Worden, H. M., Worden, J. R., Bowman, K., Kulawik, S. S., Rider, D., Osterman, G., Gunson, M., Goldman, A., Shephard, M., Clough, S. A., Rodgers, C., Lampel, M., and Chiou, L., 2006. Nadir Measurements of carbon monoxide distributions by the tropospheric emission spectrometer onboard the Aura spacecraft: overview of analysis approach and examples of initial results. *Geophysical Research Letters*, **33**, L22806, doi:10.1029/2006GL027000.
- Worden, H. M., Logan, J., Worden, J. R., Beer, R., Bowman, K., Clough, S. A., Eldering, A., Fisher, B., Gunson, M. R., Herman, R. L., Kulawik, S. S., Lampel, M. C., Luo, M., Megretskaia, I. A., Osterman, G. B., and Shephard, M. W., 2007. Comparisons of tropospheric emission spectrometer (TES) ozone profiles to ozonesondes: methods and initial results. *Journal of Geophysical Research*, **112**, D03309, doi:10.1029/2006JD007258.
- Worden, H. M., Deeter, M. N., Edwards, D. P., Gille, J. C., 2009. Multispectral retrieval of CO from MOPITT. *Eos Transactions, AGU*, 90, Fall meeting supplement, Abstract U33B–0059.
- Yang, K., Krotkov, N. A., Krueger, A. J., Carn, S. A., Bhartia, P. K., and Levelt, P. F., 2007. Retrieval of large volcanic SO₂ columns from the Aura ozone monitoring instrument: comparison and limitations. *Journal of Geophysical Research*, **112**, D24S43, doi:10.1029/2007JD008825.
- Zhang, L., et al., 2006. Ozone-CO correlations determined by the TES satellite instrument in continental outflow regions. *Geophysical Research Letters*, **33**, L18804, doi:10.1029/2006GL026399.

Cross-references

[Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties](#)
[Stratospheric Ozone](#)
[Trace Gases, Troposphere - Detection from Space](#)

ATMOSPHERIC GENERAL CIRCULATION MODELS

Joao Teixeira¹, Mark Taylor², Anders Persson³ and Georgios Matheou¹

¹Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

²Sandia National Laboratory, Albuquerque, New Mexico, USA

³United Kingdom Meteorological Office, Exeter, Devon, UK

Definition

Atmospheric general circulation models (atmospheric GCMs) are mathematical models based on numerically discretized versions of differential equations that describe the atmospheric physics and dynamics, which are utilized to simulate the global atmospheric circulation. Atmospheric GCMs have several practical applications including medium-range (typically 3–10 days) [weather forecasting](#), (see entry [Weather Prediction](#)), seasonal forecasting (typically 3–12 months) when coupled to models of other components of the climate system such as the global ocean, and climate prediction (typically 10–1,000 years) when models of the various components of the climate system, such as sea and land ice, carbon-cycle, and biosphere models, are incorporated.

Introduction and short history

Atmospheric general circulation models are based on the fundamental fluid dynamics and thermodynamics equations that govern the transport of momentum, energy, water mass, and chemical species in the atmosphere. The essential fluid dynamics and thermodynamics equations were formalized during the nineteenth century leading to the Navier–Stokes equations, the energy conservation equation, and the introduction of the Coriolis effect in the momentum conservation equations.

When meteorological weather forecast services were established around the world in the second half of the nineteenth century, their practitioners were often compared with their astronomical colleagues, who could accurately predict celestial events such as eclipses and the return of comets. However, weather forecasters, relying on empirical rules and experience, often had difficulties in accurately predicting next day's weather.

In the early twentieth century, Vilhelm Bjerknes (1904), proposed that weather prediction should be based on the fundamental equations of fluid dynamics and energy conservation. Accordingly, weather could, in principle, be predicted along the same lines as the motions of celestial bodies.

In 1922, L. F. Richardson (Richardson, 1922) publishes a groundbreaking book on numerical weather prediction where he basically sets up the agenda for the entire field and describes in detail an example of a forecast. However, and besides the fact that his forecast appeared as fairly inaccurate, it was quite clear that the amount of required

numerical computations, if done by humans, was impossibly large in operational terms so as to be able to produce forecasts much faster than the real weather. In order to overcome this problem, Richardson envisioned a future where thousands of “human calculators” would be working together in a single room producing weather forecasts for the whole world.

The situation changed dramatically with the advent of computers after World War II, when John von Neumann thought of weather prediction as one of the most obvious applications of the new tremendous numerical power provided by computers – in fact, von Neumann was not as interested in weather forecasting per se, as he was interested in computational models of the atmospheric general circulation. During this period, a substantial amount of research was carried out in creating the field of numerical analysis of differential equations. In the mid-1950s, Norman A. Phillips developed the first [general circulation model](#) of the global atmosphere (e.g., Randall, 2000).

Model dynamics

Modern global weather and climate prediction models are made up of several component models (i.e., atmosphere, land, ocean, ice). The models of the atmospheric component, or atmospheric GCMs, have two main building blocks: (1) model dynamics, representing the large-scale motions of the Earth’s atmosphere, and (2) model physics, which includes a suite of parameterizations of sub-grid-scale physical processes (see below). It is the dynamical core’s role to solve the partial differential equations that describe the large-scale motions. The heating, cooling, and mixing processes represented by the model physics appear as forcing terms in these equations. Errors in both the model physics and dynamics can have significant impacts in terms of weather and climate predictability.

Atmospheric dynamics are well approximated by numerical solutions of the Navier–Stokes equations, but as in almost all fluid dynamics applications, this approach is far too computationally expensive because of the large spectrum of scales present in global weather and climate prediction problems. Atmospheric GCMs use modified equations, which still capture important features of atmospheric dynamics, such as cyclones, anticyclones, frontal zones, and jet streams while neglecting inconsequential processes for weather and climate forecasting such as acoustic waves (see entry [Acoustic Waves, Propagation](#)). A common choice are the atmospheric *primitive equations*, which also neglect small terms related to spherical geometry, small terms related to the Coriolis effect, and some small vertical acceleration terms (the hydrostatic approximation).

The computational domain of a GCM is the thin spherical shell containing the Earth’s atmosphere. One of the defining characteristics of a dynamical core is how it handles this spherical domain. The grids used in GCMs

are constructed by decomposing a three-dimensional grid into a two-dimensional grid for conforming to the surface of the sphere (the *horizontal grid*) and a one-dimensional grid in the radial direction (the *vertical grid*). For the horizontal grid, there are three main approaches all of which have strengths and weaknesses.

One approach is to use a mesh of equally spaced points in latitude and longitude. These grids are highly structured but have a severe clustering of grid points at both poles, creating some numerical challenges collectively referred to as the pole problem. There are many effective numerical treatments of the pole problem, one of the most successful being the spherical harmonic expansions used by global spectral methods, but they often degrade parallel scalability (the ability to efficiently use multiprocessor computers in order to speed up the numerical solution of the equations).

A second approach is to decompose the sphere into several regions, each of which is meshed with a quasi-uniform grid without poles. The resulting grids must be stitched together, using a composite or overset grid method. This stitching process can make it difficult to numerically conserve mass and energy in a physically consistent way.

The third approach is to use an arbitrary tiling of the surface of the sphere with polygons, usually spherical triangles, quadrilaterals, or a combination of hexagons and pentagons. These grids are challenging for finite difference methods and are most effectively utilized by finite element and finite volume methods designed for fully unstructured grids. The choice depends on the application needs in terms of accuracy, numerical conservation properties, and parallel scalability.

Model physics

The problem of physical parameterization in fluids is as old as the first modern studies of turbulence in the nineteenth century. It was apparently clear from the start, at least to some, that for turbulent fluids, such as the atmosphere, it is not feasible or even relevant to try to follow every parcel of fluid in its turbulent trajectory. Instead, it was suggested from the early twentieth century that research should concentrate on trying to understand the statistical properties of turbulent flows.

The numerical discretizations in weather and climate prediction models imply a limit for the temporal and spatial scales below which the fluid motions cannot be resolved by the model. Because of the nonlinear nature of the atmosphere, the unresolved small scales can have a fundamental influence over the resolved large scales. Since there is no way of explicitly knowing in detail what happens at the sub-grid scales, the physics at these scales has to be parameterized as a function of the resolved motions. What can be estimated about the model variables at scales smaller than the horizontal grid scale are their statistical properties such as joint probability density functions (PDF) within the grid box. Ultimately, for

turbulence and convection, the essential problem of parameterization is how to estimate these PDFs at every model grid point.

It is natural to divide atmospheric physics parameterizations into two main groups: (1) the parameterization of radiation and cloud microphysics and (2) the parameterization of convective and turbulent mixing and associated cloud structure. Radiation and cloud microphysics parameterizations represent physical interactions that occur at extremely small scales from the atomic and molecular scales to cloud droplet scales. Turbulence and convection parameterizations, as discussed before, attempt to represent in a large-scale, atmospheric flow that occurs at scales smaller than the typical grid size (on the order of 10–100 km in the horizontal) of atmospheric GCMs. As such, they do represent fundamentally different physics and are often treated using different mathematical and conceptual tools.

Overall, in modern atmospheric GCMs, there are several different physical processes that need to be parameterized, which are traditionally classified as *follows*: turbulence or boundary layer parameterizations that represent sub-grid vertical motions within the boundary layer (i.e., the region of the atmosphere where turbulence is most prevalent, from the surface to about 1 km height); moist convection parameterizations that represent sub-grid vertical flow due to convective motions driven by condensation/evaporation processes associated with convective towers; cloud parameterizations that represent sub-grid clouds and condensation/

evaporation processes, including cloud microphysics; longwave and shortwave radiation; interactions with the ocean and land surface; and sub-grid orographic effects (e.g., gravity wave drag). More detailed information regarding atmospheric general circulation models in general, and model physics and dynamics in particular, can be found in Haltiner and Williams (1983), Randall (2000), and Washington and Parkinson (2005).

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Bjerknes, V., 1904. Das problem der wettvorhersage, betrachtet vom standpunkt der mechanik und der physic. *Meteorologische Zeitschrift*, **21**, 1–7.
- Haltiner, G. J., and Williams, R. T., 1983. *Numerical Prediction and Dynamic Meteorology*, 2nd edn. New York: Wiley, p. 477.
- Randall, D. A., 2000. *General Circulation Model Development: Past, Present, and Future*. London: Academic, Vol. 70, p. 416.
- Richardson, L. F., 1922. *Weather Prediction by Numerical Processes*. London: Cambridge University Press, p. 236.
- Washington, W., and Parkinson, C. L., 2005. *An Introduction to Three-Dimensional Climate Modeling*, 2nd edn. Sausalito: University Science, p. 354.

C

CALIBRATION AND VALIDATION

Andreas Colliander
Jet Propulsion Laboratory, California Institute of
Technology, Pasadena, CA, USA

Definition

Calibration. The process of quantitatively defining the system responses, under specified conditions, to known, controlled signal inputs. The result of a calibration permits either the assignment of values of measurands to the system output or the determination of corrections with respect to the system output (Joint Committee for Guides in Metrology JCGM (includes ISO) 2008; Randa et al., 2008; CEOS Working Group on Calibration and Validation, 2012).

Validation. The process of assessing, by independent means, the quality of the data products derived from the system outputs. The quality is determined with respect to the specified requirements (Joint Committee for Guides in Metrology JCGM (includes ISO) 2008; Randa et al., 2008; CEOS Working Group on Calibration and Validation, 2012).

Introduction

The value of remotely sensed data products, in the scientific sense in particular, is determined by how well the characteristics of a product are known (e.g., Platt and Sathyendranath, 1988; Wentz and Schabel, 2000; Atlas and Hoffman, 2000; Jung et al., 2010). These characteristics generally include long- and short-term deviation of the product value from the true value corresponding to the measurement, which is estimated through independent means (e.g., Wehr and Attema, 2001), and accuracy of the geographic location assigned to the product

(e.g., Wolfe et al., 2002; Small et al., 2004). The process of determining these characteristics for a particular remote sensing product is referred to as validation. However, before a data product is validated, it needs to be calibrated. Therefore, the calibration and validation processes are very closely linked together although they are distinctively two separate processes (see the Definition). This entry discusses calibration and validation in terms of characterization against the true value; geolocation aspect of the validation is a separate topic with specific challenges and solutions.

Remote sensing missions have requirements for the data products they are tasked to produce (e.g., Barre et al., 2008). The aim of the calibration and validation process of a particular mission is then to show that it meets its stated requirements (e.g., Delwart et al., 2008). Since the requirements are typically assigned based on expected scientific utilization of the data, the calibration and validation processes are generally regarded as a scientific function. Furthermore, the science community commonly contributes to calibration and validation efforts of data products independently from the missions in their research, due to the importance of knowing the characteristics and quality of the data (e.g., Donlon et al., 2002; Wang and Key, 2003; Mears and Wentz, 2005; Flanner et al., 2010).

The challenges of calibration and validation are specific to the mission and the data product. However, there are some general challenges concerning most of the remote sensing products. The most common and general issues causing concern are (1) establishment of accurate reference sites where the true value corresponding to the measurement can be estimated independently and accurately (e.g., Cosh et al., 2004) and (2) representing the entire measurement domain, which is often global, with a finite number of these sites (e.g., Morisette et al., 2002). It is typical the calibration and validation effort of a given mission

continues during the entire lifetime of a mission and even beyond (e.g., Xu and Ignatov, 2010).

Remote sensing data products can be divided into sensor products and geophysical products. Sensor product refers to output of an instrument after translating the instrument counts to a desired quantity, such as normalized radar cross section (e.g., Srivastava et al., 1999) or radiance (e.g., Abrams, 2000). Geophysical products refer to data products which contain geophysical parameters, such as wind speed (e.g., Liu et al., 1998) or leaf area index (e.g., Yang et al., 2006), retrieved based on the sensor products (and usually with some additional ancillary data). The calibration and validation of the sensor and geophysical products differ in some aspects and are discussed separately in the subsequent text.

International cooperation is in a key role to satisfy the requirements of calibration and validation of typical remote sensing products with very large domains. Committee on Earth Observation Satellites (CEOS) (CEOS Working Group on Calibration and Validation, 2012) is one international organization which has been active in promoting calibration and validation efforts. The Working Group on Calibration and Validation of CEOS has formulated a general approach for calibration and validation of remote sensing products and has established a validation hierarchy based on different stages of extent of validation efforts (see the end of this text).

Historical perspective

At the beginning of the satellite remote sensing era (e.g., Nimbus-1 in 1964 and Landsat-1 in 1972), the calibration and validation activities were mostly limited to activities carried out directly by the space agencies. The current form of utilization of remote sensing data products started roughly with the launch of NASA's Nimbus-7 satellite in 1978. The new data policy of this mission enabled engagement of wider science community in more rapid manner after the launch of the satellite (Goddard Space Flight Center, National Aeronautics and Space Administration 2004). This also contributed to the start of the community-wide pre-and postlaunch calibration and validation efforts of remote sensing data products (e.g., Austin, 1980; Hovis, 1982; Stowe, 1982; Bernstein and Chelton, 1985; Stowe et al., 1988). After this, most NASA Earth observation satellites have followed the similar data policy. However, other space agencies have highly varying policies regarding data dissemination, which directly affects the extent of the calibration and validation activity.

Currently, it is common that a launch of each new remote sensing instrument initiates the science community to seek opportunities to participate in the calibration and validation activities. Naturally, the increased number of instruments and accumulated experience on the calibration and validation of satellite data products is another main reason for the increasing activity in the calibration and validation front. However, the calibration and validation efforts of remote sensing products have maintained

very similar features from the earlier days of remote sensing (e.g., compare (Hilland et al., 1985) with (O'Carroll et al., 2008)). At the same time, new instruments and new applications do require new methods for successful calibration and validation of remote sensing products.

Sensor products

Calibration and validation of the sensor products of a mission is the critical part in ensuring the usefulness of the mission data. The quality of the sensor products typically dominates the quality of the geophysical products. Each remote sensing instrument has an algorithm which is used to translate the raw instrument counts to the desired quantity. The complexity of the algorithm depends on the instrument implementation and the properties of the desired quantity. For example, retrieval of normalized radar cross section requires measurement geometry in addition to instrument parameters (Ulaby et al., 1982), whereas antenna temperature of radiometer is independent of the measurement geometry (Ulaby et al., 1981). In principle, the features of the algorithm dictate the requirements for the calibration effort (i.e., parameters to be adjusted), and the quantity itself determines the requirements for the validation effort (i.e., proper target representing the quantity).

Instrument calibration usually includes some sort of internal calibration sources (e.g., Xiong and Barnes, 2006; Brown et al., 2007). While these sources can be used to remove the effects of some instrument non-idealities, they do not provide reference for the full instrument measurement chain (e.g., Butler and Barnes, 1998). There are different approaches for external calibration: use of an onboard reference target (e.g., Yamaguchi et al., 1998; Twarog et al., 2006); measurement of celestial targets, such as moon or cosmic microwave background radiation (e.g., Sun et al., 2003; Jones et al., 2006); or establishment of a reference target on the ground. Dedicated efforts to improve the stability of the observations (e.g., Gopalan et al., 2009; Eymard et al., 2005) and studies to correct errors caused by the antenna of an instrument (e.g., Njoku, 1980; McKague et al., 2011) are also typical for calibration of remote sensing instrument. Intercalibration between the remote sensing instruments is an important aspect for extending a data record either in time, to lengthen the time series and/or increase the fidelity of the time series, or in space to increase coverage (e.g., Cavalieri et al., 2012; Xiong et al., 2008).

There are areas on the surface of the Earth which provide well-defined response for some types of remote sensing measurements. Therefore, the target on the ground may be a natural scene suitable for calibration purposes (e.g., rain forests for microwave scatterometers (Long and Skouson, 1996)) or it can also be a target built for this specific reason (e.g., corner reflector for synthetic aperture radar (Shimada et al., 2009)). Some of the natural targets can provide relative well-defined absolute reference

value; others are more suitable for just tracking stability of the instrument. Examples of Earth scenes used as vicarious references for spaceborne remote sensing instrument calibration are Antarctica ice sheets (e.g., Macelloni et al., 2007, 2011), Amazon rain forests (e.g., Shimada, 2005), oceans (Ruf et al., 2006), deserts (e.g., Slater et al., 1987), and dry lake beds (e.g., Helder et al., 2010). Utility of man-made structures has also been demonstrated, such as a large asphalt field in Biggar et al., (2003).

Validation of sensor data products is usually done by using on-ground reference targets discussed above, since this represents the relevant measurement plane for the scientific utilization of the measurements. After the sensor product has been calibrated, it is compared against selected targets to establish the uncertainty of the product. This process may lead to further calibration too in which case the residual deviation from the targets becomes the result of the validation. Ideally, of course, the targets used for calibration and validation should be different. However, the use of the same targets is the reason why sometimes the nomenclature of sensor cal/val process refers only to calibration and does not include references to validation, even though it is clearly part of the process.

Geophysical products

The retrieval algorithms of geophysical parameters are highly varying in their approach to determine the value of the desired parameter. Regardless of the approach, however, each algorithm requires calibration in order to optimize the correctness of its output. In order for the calibration process to be successful, the structure and error contributions of the algorithm need to be known (e.g., Pulliainen et al., 1993; Keihm et al., 1995; Wentz, 1997; Njoku et al., 2003; Brando and Dekker, 2003). Several algorithms include forward models that require detailed calibration before application to the inverse processing (e.g., Wigneron et al., 2007). In order to accomplish the calibration of an algorithm field, measurements and additional remote sensing measurements are typically exploited to determine the parameter values of the algorithm (e.g., Kelly et al., 2003; Njoku et al., 2003).

The validation process of geophysical products requires knowledge of the true value of the geophysical parameter within the effective measurement area with uncertainty less than the required uncertainty of the product. The following subsection discusses the issues related to scaling the in situ truth measurement to the footprint scale in spatial domain. Even in the absence of the spatial scaling issues, the uncertainty of the actual in situ measurement must be less than the uncertainty requirement of the product (e.g., Emery et al., 2001; Bailey and Werdell, 2006; Henocq et al., 2010). The establishment of these validation sites depends naturally a great deal on the geophysical parameter: The general approach and requirements for wind speed (e.g., Dobson et al., 1987) or chlorophyll a (e.g., Ruiz-Verdu et al., 2008) measurement are completely different from snow water equivalent

(e.g., Tedesco and Narvekar, 2010) or leaf area index (e.g., Garrigues et al., 2008) measurements, let alone atmospheric water vapor (e.g., Divakarla et al., 2006), or ozone (e.g., Froidevaux et al., 2008) measurements. The objective in each case is nevertheless the same, to find a representative measurement of the parameter so that it can be compared against the remotely sensed value. After appropriately matching up the remotely sensed product and the in situ measurement, the validation results are typically presented as, for example, root mean square error, correlation, and histograms (e.g., Hooker and McClain, 2000; Bourassa et al., 2003; Hilland et al., 1985).

Spatial scaling

Remote sensing measurements are based on the instrument recordings of interaction of electromagnetic waves with the target. The instruments have a defined sensing area or volume (i.e., footprint) depending on the antenna beam shape and interaction of the measurement signal with the sensed medium. When it comes to calibrating and validating the measurements, the independent reference measurements, in situ measurements in particular, typically do not have the same features as the remotely sensed signal and do not measure the exactly same domain as the footprint represents. The translation of the reference measurements to the remote sensing footprint is often referred to as spatial scaling, and it is a crucial part of the calibration and validation of remote sensing products. Sensor product calibration and validation efforts usually try to utilize homogeneous regions where scaling is not an issue in the same way as with calibration and validation of typical geophysical products.

The challenge of the spatial scaling depends typically on the relationship between the heterogeneity of the measured parameter and the size of the footprint. Scaling of even relative high-resolution (small size) footprint may be challenging for highly heterogeneous parameters (e.g., Liang et al., 2002). Some remote sensing instruments have very low resolution, but if the measured parameter changes slowly over large distances, scaling can be accomplished with relative few resources within the footprint (e.g., Le Vine et al., 2007). The most challenging cases include of course remote sensing measurements of highly heterogeneous parameters with large footprints (e.g., Jackson et al., 2010).

There are several techniques developed for scaling the value of geophysical parameters up to the footprints of remote sensing measurements. As an example, these techniques include aggregation of in situ measurements (e.g., Jackson et al., 2010), model-based techniques (e.g., Chen et al., 1999), timing of the acquisition so that the heterogeneity effect is minimized (e.g., Wang et al., 2008), and temporal stability approach which assumes that single point of the area represents the footprint average (e.g., the challenge then is to find the representative point (Vachaud et al., 1985; Grayson and

Western, 1998)). An important aspect of the upscaling of the in situ measurements is estimation of the error associated with the upscaled value. Several techniques have been proposed and used for accomplishing this, for example: investigating the variance of the subscale measurements (e.g., Tian et al., 2002) and combination of several observation sources (e.g., Hilland et al., 1985; O'Carroll et al., 2008; Caires and Sterl, 2003; Miralles et al., 2010).

Coverage

Although it is generally accepted that the validation of the geophysical products should be done against in situ measurements, other references are being applied too. The reason for this is that in general, in situ measurements have limited coverage in space and in time, i.e., they do not cover the entire domain, which is often global (the main reason why remote sensing is applied in the first place), and it may not always be possible to make in situ measurements for long periods or with high frequency for a certain location (e.g., consider limitations of radiosondes, dropsondes, and buoys).

Depending on the size of the covered domain, the challenge for the calibration and validation effort is to find a strategy with which it can be claimed that the product is calibrated and validated over the entire domain. Therefore, other remote sensing sources (e.g., Corlett et al., 2006) and models (e.g., Caires and Sterl, 2003) have been used to complement the in situ measurements. It is also typical to divide the entire domain in sub-domains and then find validation sites which represent each sub-domain (e.g., Hilland et al., 1985) or to cover a diversity of conditions with set of sites which can be claimed to represent the conditions over the entire domain (e.g., Ceccato et al., 2002). But only rarely a validation is accepted without some strategy to reference the measurements back to verifiable in situ acquisitions.

Temporal context

Remote sensing products can be calibrated and validated over a short time period (a few years at most) or a long time period (at least a decade). Calibration and validation of individual missions is usually a short-term effort because the mission requirements are typically set without long-term requirement and, furthermore, the duration of a single mission is seldom long enough to qualify as a long term anyway. Therefore, long-term calibration and validation imply inter-mission effort to extend the calibrated and validated data record to over decade long time frame (e.g., Gallo et al., 2005). Typically high quality long-term data records are required for climate applications (e.g., Flanner et al., 2011; Behrenfield et al., 2006), but also other monitoring and tracking applications require long-term remote sensing observations (e.g., Lepers et al., 2005). The type, quality, and length of remote sensing data records have limited the use of remote sensing data for long-term applications until more recently.

The requirements of calibration and validation process are affected significantly by the climatic temporal context of multiple decades combined with usually very high requirements on the stability. Therefore, a combination of both retrieval intercomparisons and in situ measurements is often necessary to validate the long-term record (e.g., Takala et al., 2009). For example, the importance of [sea surface temperature \(SST\)](#) record for climate studies was understood long time ago (e.g., Harries et al., 1983) and ever since significant effort has been made to establish a well-calibrated long-term SST record (e.g., Stuart-Menteth et al., 2003).

Prelaunch calibration and validation

It is common that remote sensing missions include calibration and validation activities for both sensor and geophysical retrieval algorithms in the prelaunch phase. Essentially, the objective of these activities is to increase the expectation of the mission success to the level that launching a satellite seems worthwhile. Missions set requirements for the instrument performance based on the intended use of the measurements. In the prelaunch phase, this performance is verified through measurements, analysis, and simulations. The calibration strategy of an instrument may require measurement of certain calibration parameters on the ground, and sometimes, when possible, the instrument is entirely calibrated on the ground already. These activities are often referred to as prelaunch calibration activities. However, it should be emphasized that the eventually applicable calibration is almost always conducted on the orbit.

The development of geophysical retrieval algorithms starts before mission definition in research activities which try to identify potential remote sensing measurements concepts. The prelaunch calibration and validation activities include similar components as retrieval algorithm research such as field campaigns and simulations. However, these activities are driven by particular mission characteristics such as exact observation configuration including measurement frequency, footprint, coverage, and instrument performance figures. The prelaunch efforts can only approximate the actual measurements of the mission, and the actual calibration and validation of the retrieval algorithms and products takes place only after the launch of the mission.

Validation stages

CEOS (CEOS Working Group on Calibration and Validation, 2012) has put forward a four-stage validation hierarchy which has been adopted by many data providers. The validation stage increases with increasing product maturity and extensiveness of the validation effort.

- Stage 1 validation: Product accuracy is assessed from a small (typically <30) set of locations and time periods by comparison with in situ or other suitable reference data.

- Stage 2 validation: Product accuracy is estimated over a significant set of locations and time periods by comparison with reference in situ or other suitable reference data. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and time periods. Results are published in the peer-reviewed literature.
- Stage 3 validation: Uncertainties in the product and its associated structure are well quantified from comparison with reference in situ or other suitable reference data. Uncertainties are characterized in a statistically robust way over multiple locations and time periods representing global conditions. Spatial and temporal consistency of the product and with similar products has been evaluated over globally representative locations and periods. Results are published in the peer-reviewed literature.
- Stage 4 validation: Validation results for stage 3 are systematically updated when new product versions are released and as the time series expands.

Summary

All remote sensing products require calibration and validation, and it is an essential part of the process of making remote sensing products to meet the requirements of scientific utilization. The main challenges in the calibration and validation of almost any data product are how to make corresponding and representative reference measurements and how to extend the validation over the entire measurement domain. These challenges are overcome by careful design of the reference sites and their in situ measurements, which vary greatly depending on the geophysical parameter and on the footprint of the remote sensing product, and strategizing the utilization of diversity of validation sites with augmentation by other remote sensing products and models. Intersensor and inter-product calibration and validation are conducted to increase the length and fidelity of the time series or the coverage of the measurement. This is essential for utilization of remote sensing data for climate change studies.

Acknowledgments

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Abrams, M., 2000. The advanced spaceborne thermal emission reflection radiometer (ASTER): data products for the high spatial resolution imager on NASA's terra platform. *International Journal of Remote Sensing*, **21**(5), 847–859.
- Atlas, R., and Hoffman, R. N., 2000. The use of satellite surface wind data to improve weather analysis and forecasting at the NASA data assimilation office. *Elsevier Oceanography Series*, **63**, 57–78.
- Austin, R. W., 1980. Gulf of Mexico, ocean-color surface-truth measurements. *Boundary-Layer Meteorology*, **18**(3), 269–285.

- Bailey, S. W., and Werdell, J. P., 2006. A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment*, **102**(1–2), 12–23.
- Barre, H. M. J. P., Duesmann, B., and Kerr, Y. H., 2008. SMOS: the mission and the system. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(3), 587–593.
- Behrenfield, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., Miligan, A. J., Falkowski, P. G., Letelier, R. M., and Boss, E. S., 2006. Climate-driven trends in contemporary ocean productivity. *Nature*, **444**, 752–755.
- Bernstein, R. L., and Chelton, D. B., 1985. Large-scale sea surface temperature variability from satellite and shipboard measurements. *Journal of Geophysical Research*, **90**(C6), 11619–11630.
- Biggar, S. F., Thome, K. J., and Wisniewski, W., 2003. Vicarious radiometric calibration of eo-1 sensors by reference to high-reflectance ground targets. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6), 1174–1179.
- Bourassa, M. A., Legler, D. M., O'Brien, J. J., and Smith, S. R., 2003. Seawinds validation with research vessels. *Journal of Geophysical Research*, **108**, C2.
- Brando, V. E., and Dekker, A. G., 2003. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6), 1378–1387.
- Brown, S. T., Desai, S., Lu, W., and Tanner, A. B., 2007. On the long-term stability of microwave radiometers using noise diodes for calibration. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(7), 1908–1920.
- Butler, J. J., and Barnes, R. A., 1998. Calibration strategy for the earth observing system (EOS) – AM1 platform. *IEEE Transactions on Geoscience and Remote Sensing*, **36**(4), 1056–1061.
- Caires, S., and Sterl, A., 2003. Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical Research*, **108**, C3.
- Cavalieri, D. J., Parkinson, C. L., DiGirolamo, N., and Ivanoff, A., 2012. Intersensor calibration between F13 SSMI and F17 SSMIS for global sea ice data records. *IEEE Geoscience and Remote Sensing Letters*, **9**(2), 233–236.
- Ceccato, P., Flasse, S., and Gregoire, J.-M., 2002. Designing a spectral index to estimate vegetation water content from remote sensing data: part 2. Validation and applications. *Remote Sensing of Environment*, **82**(2–3), 198–207.
- CEOS Working Group on Calibration and Validation, 2012. Committee on Earth Observation Satellites (Online). Available from <http://www.ceos.org>. Accessed 14 Feb 2012.
- Chen, J. M., Liu, J., Cihlar, J., and Goulden, M. L., 1999. Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications. *Ecological Modeling*, **124**(2–3), 99–119.
- Corlett, G. K., et al., 2006. The accuracy of SST retrievals from AATSR: an initial assessment through geophysical validation against in situ radiometers, buoys and other SST data sets. *Advances in Space Research*, **37**(4), 764–769.
- Cosh, M. H., Jackson, T. J., Bindlish, R., and Prueger, J. H., 2004. Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. *Remote Sensing of Environment*, **92**, 427–435.
- Delwart, S., Bouzinac, C., Wursteisen, P., Berger, M., Drinkwater, M., Martin-Neira, M., and Kerr, Y. H., 2008. SMOS validation and the COSMOS campaigns. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(3), 695–704.
- Divakarla, M. G., Barnett, C. D., Goldberg, M. D., McMillin, L. M., Maddy, E., Wolf, W., Zhou, L., and Liu, X., 2006. Validation of atmospheric infrared sounder temperature and water vapour retrievals with matched radiosonde measurements and forecasts. *Journal of Geophysical Research*, **111**, D09S15.

- Dobson, E., Monaldo, F., Goldhirsh, J., and Wilkerson, J., 1987. Validation of geosat altimeter-derived wind speeds and significant wave heights using buoy data. *Journal of Geophysical Research*, **92**(C10), 10719–10731.
- Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., and Murray, M. J., 2002. Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, **15**(4), 353–369.
- Emery, W. J., Baldwin, D. J., Schluskel, P., and Reynolds, R. W., 2001. Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *Journal of Geophysical Research*, **106**(C2), 2387–2405.
- Eymard, L., Obligis, E., Tran, N., Karbou, F., and Dedieu, M., 2005. Long-term stability of ERS-2 and TOPEX microwave radiometer in-flight calibration. *IEEE Transactions on Geoscience and Remote Sensing*, **43**(5), 1144–1158.
- Flanner, M. G., Shell, K. M., Barlage, M., Perovich, D. K., and Tschudi, M. A., 2010. Radiative forcing and albedo feedback from the northern hemisphere cryosphere between 1979 to 2008. *Nature Geoscience*, **4**, 151–155.
- Flanner, M. G., Shell, K. M., Barlage, M., Perovich, D. K., and Tschudi, M. A., 2011. Radiative forcing and albedo feedback from the northern hemisphere cryosphere between 1979 and 2008. *Nature Geoscience*, **4**, 151–155.
- Froidevaux, L., et al., 2008. Validation of aura microwave limb sounder stratospheric ozone measurements. *Journal of Geophysical Research*, **113**, D15S20.
- Gallo, K., Ji, L., Reed, B., Eidenshink, J., and Dwyer, J., 2005. Multi-platform comparisons of MODIS and AVHRR normalized difference vegetation index data. *Remote Sensing of Environment*, **99**(3), 221–231.
- Garrigues, S., Lacaze, R., Baret, F., Morisette, J. S., Weiss, M., Nickeson, J. E., Fernandes, R., Plummer, S., Shabanov, N. V., Myneni, R. B., Knyazikhin, Y., and Yang, W., 2008. Validation and intercomparison of global leaf area index products derived from remote sensing data. *Journal of Geophysical Research*, **113**, G02028.
- Goddard Space Flight Center, National Aeronautics and Space Administration, 2004. *Nimbus Program History, National Aeronautics and Space Administration*, Greenbelt: MD.
- Gopalan, K., Jones, W. L., Biswas, S., Bilanow, S., Wilheit, T., and Kasparis, T., 2009. A time-varying radiometric bias correction for the TRMM microwave imager. *IEEE Transactions on Geoscience and Remote Sensing*, **47**(11), 3722–3730.
- Grayson, R. B., and Western, A. W., 1998. Towards areal estimation of soil water content from point measurements: time and space stability of mean response. *Journal of Hydrology*, **207**, 68–82.
- Harries, J. E., Llewellyn-Jones, D. T., Minnett, P. J., Saunders, R. W., and Zavody, A. M., 1983. Observations of sea-surface temperature for climate research. *Philosophical Transactions of the Royal Society of London A*, **309**, 381–395.
- Helder, D. L., Basnet, B., and Morstad, D. L., 2010. Optimized identification of worldwide radiometric pseudo-invariant calibration sites. *Canadian Journal of Remote Sensing*, **36**(5), 527–539.
- Henocq, C., Boutin, J., Petitcolin, F., Reverdin, G., Arnault, S., and Lattes, P., 2010. Vertical variability of near-surface salinity in the tropics: consequences for L-band radiometer calibration and validation. *Journal of Atmospheric and Oceanic Technology*, **27**, 192–209.
- Hilland, J. E., Chelton, D. B., and Njoku, E. G., 1985. Production of global sea surface temperature fields for the jet propulsion laboratory workshop comparisons. *Journal of Geophysical Research*, **90**(C6), 11642–11650.
- Hooker, S. B., and McClain, C. R., 2000. The calibration and validation of seaWiFS data. *Progress in Oceanography*, **45**, 427–465.
- Hovis, W., 1982. Results of the CZCS validation program. In *OCEANS 82*. Washington.
- Jackson, T. J., Cosh, M. H., Bindlish, R., Starks, P. J., Bosch, D. D., Seyfried, M., Goodrich, D. C., Moran, S., and Du, J., 2010. Validation of advanced microwave scanning radiometer soil moisture products. *IEEE Transactions on Geoscience and Remote Sensing*, **48**(12), 4256–4272.
- Joint Committee for Guides in Metrology JCGM (includes ISO, Geneva, Switzerland), 2008. *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM)*.
- Jones, W. L., Park, J. D., Soisuvarn, S., Hong, L., Gaiser, P. W., and St. Germain, K. M., 2006. Deep-space calibration of the windsat radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(3), 476–495.
- Jung, M., et al., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467**, 951–954.
- Keihm, S. J., Janssen, M. A., and Ruf, C. S., 1995. TOPEX/poseidon microwave radiometer (TMR): III. Wet troposphere range correction algorithm and pre-launch error budget. *IEEE Transactions on Geoscience and Remote Sensing*, **33**(1), 147–161.
- Kelly, R. E., Chang, A. T., Tsang, L., and Foster, J. L., 2003. A prototype AMSR-E global snow area and snow depth algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(2), 230–242.
- Le Vine, D. M., Lagerloef, G. S. E., Colomb, F. R., Yueh, S. H., and Pellerano, F. A., 2007. Aquarius: an instrument to monitor surface salinity from space. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(7), 2040–2050.
- Lepers, E., Lambin, E. F., Janetos, A. C., DeFries, R., Achard, F., Ramankutty, N., and Scholes, R. J., 2005. A synthesis of information on rapid land-cover change for the period 1981–2000. *BioScience*, **55**(2), 115–124.
- Liang, S., Fang, H., Chen, M., Shuey, C. J., Walthall, C., Daughtry, C., Morisette, J., Schaaf, C., and Strahler, A., 2002. Validating MODIS land surface reflectance and albedo products: methods and preliminary results. *Remote Sensing of Environment*, **83** (1–2), 149–162.
- Liu, T. W., Tang, W., and Polito, P. S., 1998. NASA scatterometer provides global ocean-surface wind fields with more structures than numerical weather prediction. *Geophysical Research Letters*, **25**(6), 761–764.
- Long, D. G., and Skouson, G. B., 1996. Calibration of spaceborne scatterometers using tropical rain forests. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(2), 413–424.
- Macelloni, G., Brogioni, M., Pampaloni, P., and Cagnati, A., 2007. Multifrequency microwave emission from the dome-C area on the east Antarctic plateau: temporal and spatial variability. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(7), 2029–2039.
- Macelloni, G., Brogioni, M., Pettinato, S., Zasso, R., Crepaz, A., Zaccaria, J., Padovan, B., Drinkwater, M., 2011. Multifrequency microwave emission of the east Antarctic plateau. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2011, Vancouver: Canada.
- McKague, D. S., Ruf, C. S., and Puckett, J. J., 2011. Beam spilling correction for spaceborne microwave radiometers using the two-point vicarious calibration method. *IEEE Transactions on Geoscience and Remote Sensing*, **49**(1), 21–27.
- Mears, C. A., and Wentz, F. J., 2005. The effect of diurnal correction on satellite-derived lower tropospheric temperature. *Science*, **309**(5740), 1548–1551.
- Miralles, D. G., Crow, W. T., and Cosh, M. H., 2010. Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *Journal of Hydrometeorology*, **11**, 1423–1429.

- Morisette, J. T., Privette, J. L., and Justice, C. O., 2002. A framework for the validation of MODIS land products. *Remote Sensing of Environment*, **83**(1–2), 77–96.
- Njoku, E. G., 1980. Antenna pattern correction procedures for the scanning multichannel microwave radiometer (SMMR). *Boundary-Layer Meteorology*, **18**, 79–98.
- Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, S. T. K., and Nghiem, S. V., 2003. Soil moisture retrieval from AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(2), 215–229.
- O’Carroll, A. G., Eyre, J. R., and Saunders, R. W., 2008. Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *Journal of Atmospheric and Oceanic Technology*, **25**, 1197–1207.
- Platt, T., and Sathyendranath, S., 1988. Oceanic primary production: estimation by remote sensing at local and regional scales. *Science*, **241**, 1613–1620.
- Pulliainen, J., Karna, J.-P., and Hallikainen, M., 1993. Development of geophysical retrieval algorithms for the MIMR. *IEEE Transactions on Geoscience and Remote Sensing*, **31**(1), 268–277.
- Randa, J., Lahtinen, J., Camps, A., Gasiewski, A. J., Hallikainen, M., Le Vine, D. M., Martin-Neira, M., Piepmeier, J., Rosenkranz, P. W., Ruf, C. S., Shiue, J., Skou, N., 2008. *Recommended Terminology for Microwave Radiometry*, National Institute of Standards and Technology. Technical Note 1551.
- Ruf, C. S., Hu, Y., and Brown, S. T., 2006. Calibration of windsat polarimetric channels with a vicarious cold reference. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(3), 470–475.
- Ruiz-Verdu, A., Koponen, S., Heege, T., Doerffer, R., Brockmann, C., Kallio, K., Pyhalhti, T., Pena, R., Polvorinos, A., Heblinski, J., Ylostalo, P., Conde, L., Odermatt, D., Estelles, V., Pulliainen, J., 2008. Development of MERIS lake water algorithms: validation results from Europe. In *Proceedings of 2nd MERIS User Workshop*, Frascati: Italy.
- Shimada, M., 2005. Long-term stability of L-band normalized radar cross section of amazon rainforest using the JERS-1 SAR. *Canadian Journal of Remote Sensing*, **31**(1), 132–137.
- Shimada, M., Isoguchi, O., Tadono, T., and Isono, K., 2009. PALSAR radiometric and geometric calibration. *IEEE Transactions on Geoscience and Remote Sensing*, **47**(12), 3915–3932.
- Slater, P. N., Biggar, S. F., Holm, R. G., Jackson, R. D., Mao, Y., Moran, M. S., Palmer, J. M., and Yuan, B., 1987. Reflectance- and radiance-based methods for the in-flight absolute calibration of multispectral sensor. *Remote Sensing of Environment*, **22**(1), 11–37.
- Small, D., Rosich, B., Meier, E., Uesch, D., 2004. Geometric calibration and validation of ASAR imagery. In *Proceedings of CEOS SAR Workshop*, Ulm: Germany.
- Srivastava, S. K., Hawkins, R. K., Lukowski, T. I., Banik, B. T., Adamovic, M., and Jefferies, W. C., 1999. RADARSAT image quality and calibration – update. *Advances in Space Research*, **23**(8), 1487–1496.
- Stowe, L. L., 1982. Validation of nimbus-7 temperature-humidity infrared radiometer estimates of cloud type and amount. *Advances in Space Research*, **2**(6), 15–19.
- Stowe, L. L., Wellemeyer, C. G., Yeh, H. Y. M., Eck, T. F., The Nimbus-7 Cloud Data Processing Team, 1988. Nimbus-7 global cloud climatology. Part I: algorithms and validation. *Journal of Climate*, **1**(5), 445–470.
- Stuart-Menteth, A. C., Robinson, I. S., and Challenor, P. G., 2003. A global study of diurnal warming using satellite-derived sea surface temperature. *Journal of Geophysical Research*, **108**, C5.
- Sun, J., Xiong, X., Guenther, B., and Barnes, W., 2003. Radiometric stability monitoring of the MODIS reflective solar bands using the moon. *Metrologia*, **40**, S85–S88.
- Takala, M., Pulliainen, J., Metsamaki, S. J., and Koskinen, J. T., 2009. Detection of snowmelt using spaceborne microwave radiometer data in eurasia from 1979 to 2007. *IEEE Transactions on Geoscience and Remote Sensing*, **47**(9), 2996–3007.
- Tedesco, M., and Narvekar, P. S., 2010. Assessment of the NASA AMSR-E SWE product. *IEEE Transactions on Geoscience and Remote Sensing*, **3**(1), 141–159.
- Tian, Y., et al., 2002. Multiscale analysis and validation of the MODIS LAI product I. Uncertainty assessment. *Remote Sensing of Environment*, **83**, 414–430.
- Twarog, E. M., Purdy, W. E., Gaiser, P. W., Cheung, K. H., and Kelm, B. E., 2006. WindSat on-orbit warm load calibration. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(3), 516–529.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing: Active and Passive*. Norwood: Artech House. Microwave Remote Sensing Fundamentals and Radiometry, Vol. 1.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1982. *Microwave Remote Sensing: Active and Passive*. Reading: Addison-Wesley Company. Radar Remote Sensing and Surface Scattering and Emission Theory, Vol. 2.
- Vachaud, G., Passerat de Silans, A., Balabanis, P., and Vauclin, M., 1985. Temporal stability of spatially measured soil water probability density function. *Journal of Soil Science Society of America*, **49**(4), 822–828.
- Wang, X., and Key, J. R., 2003. Recent trends in arctic surface, cloud, and radiation properties from space. *Science*, **299**(5613), 1725–1728.
- Wang, W., Liang, S., and Meyers, T., 2008. Validating MODIS land surface temperature products using long-term nighttime ground measurement. *Remote Sensing of Environment*, **112**(3), 623–635.
- Wehr, T., and Attema, E., 2001. Geophysical validation of envisat data products. *Advances in Space Research*, **28**(1), 83–91.
- Wentz, F. J., 1997. A well-calibrated ocean algorithm for special sensor microwave/imager. *Journal of Geophysical Research*, **102**(C4), 8703–8718.
- Wentz, F. J., and Schabel, M., 2000. Precise climate monitoring using complementary satellite data sets. *Nature*, **403**, 414–416.
- Wigneron, J.-P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., Ferrazzoli, P., de Rosnay, P., Gurney, R., Calvet, J.-C., Grant, J. P., Guglielmetti, M., Hornbuckle, B., Matzler, C., Pellarin, T., and Schwank, M., 2007. L-band microwave emission of the biosphere (L-MEB) model: description and calibration against experimental data sets over crop fields. *Remote Sensing of Environment*, **107**(4), 639–655.
- Wolfe, R. E., Nishihama, M., Fleig, A. J., Kuypier, J. A., Roy, D. P., Storey, J. C., and Patt, F. S., 2002. Achieving sub-pixel geolocation accuracy in support of MODIS land science. *Remote Sensing of Environment*, **83**(1–2), 31–49.
- Xiong, X., and Barnes, W., 2006. An overview of MODIS radiometric calibration and characterization. *Advances in Atmospheric Sciences*, **23**(1), 69–79.
- Xiong, X., Sun, J., and Barnes, W., 2008. Intercomparison of on-orbit calibration consistency between terra and aqua MODIS reflective solar bands using the moon. *IEEE Geoscience and Remote Sensing Letters*, **5**(4), 778–782.
- Xu, F., and Ignatov, A., 2010. Evaluation of in situ surface temperature for use in the calibration and validation of satellite retrievals. *Journal of Geophysical Research*, **115**, C09022.
- Yamaguchi, Y., Kahle, A. B., Tsu, H., Kawakami, T., and Pniel, M., 1998. Overview of advanced spaceborne thermal emission and reflection radiometer (ASTER). *IEEE Transactions on Geoscience and Remote Sensing*, **36**(4), 1062–1071.

Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., Privette, J. L., Huemmrich, K. F., Fensholt, R., Sandholt, I., Weiss, M., Ahl, D. E., Gower, S. T., Nemani, R. R., Knyazikhin, Y., and Myneni, R. B., 2006. MODIS leaf area index products: from validation to algorithm improvement. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(7), 1885–1898.

CALIBRATION, MICROWAVE RADIOMETERS

Christopher Ruf

Department of Atmospheric, Oceanic and Space Sciences,
University of Michigan, Ann Arbor, MI, USA

Definition

Antenna temperature (T_a). A weighted average of the brightness temperature incident on a radiometer from all directions, where the weighting is the antenna radiation pattern.

Receiver noise temperature (T_{rec}). The equivalent brightness temperature that would produce the same signal strength as the noise generated by a radiometer's electronics.

System noise temperature (T_{sys}). The sum of the antenna temperature entering a radiometer plus its receiver noise temperature.

Antenna radiation pattern. The sensitivity of an antenna to incoming radiation, usually expressed as a function of angle of arrival in polar coordinates and normalized so that the integral over all angles is unity.

Antenna mainbeam. The portion of an antenna radiation pattern localized near the direction of maximum sensitivity.

Antenna sidelobes. The antenna radiation pattern away from its mainbeam.

Overview

A microwave radiometer measures the power associated with a particular polarization component of a propagating electromagnetic wave over a specific portion of the electromagnetic spectrum (see entry *Radiation, Electromagnetic*). The brightness temperature (T_b) in the direction from which the wave originated can be inferred from the power measurement. The T_b is specific to the polarization state and spectral range of the measurement. Raw radiometer power measurements generally contain contributions from other sources than the desired T_b , including incoming radiation from other directions and polarizations as well as radiation emitted by the radiometer receiver itself. Radiometer calibration consists of an accurate accounting for all significant contributions to the raw measurements and estimation of the desired T_b .

Antenna effects

In general, electromagnetic radiation incident on an antenna arrives from all directions and with all possible polarization states. An ideal antenna would selectively receive radiation with only the desired polarization and that arrives only from the desired direction. In practice, an antenna will receive radiation incident from all directions weighted by its radiation pattern. The weighted average of the incident T_b , weighted by the radiation pattern, is referred to as the antenna temperature (T_a). The mainbeam of the radiation pattern is the (typically narrow) range of directions over which the antenna is most sensitive. Calibrated T_b measurements by a radiometer are generally considered to originate from this range of directions. The sidelobes of the radiation pattern describe its sensitivity to radiation incident from all other directions outside the mainbeam. Typically 1–10% of the total power measured by a radiometer originates from sidelobe directions. Similarly, an antenna will receive radiation from all polarization states, weighted by its sensitivity to each component. An antenna's polarimetric gain matrix describes the sensitivity. Its main diagonal elements specify the sensitivity to the desired polarization, in a manner analogous to the mainbeam for directional sensitivity. The off-diagonal elements play a similar role as the antenna sidelobes, specifying the sensitivity to unwanted polarization states.

Calibration of a radiometer for antenna effects consists of correcting the T_a measurements for each of the unwanted contributors in order to estimate the T_b contribution with the desired polarization arriving from the mainbeam direction. Sidelobe corrections are generally made by (1) measuring or modeling the radiation pattern of the antenna; (2) estimating the mean T_b entering the antenna from its sidelobes; and then (3) subtracting the mean sidelobe T_b , weighted by the sidelobe sensitivity, from the raw measurements (Njoku et al., 1980). Similarly, polarimetric corrections first require that an antenna's gain matrix be known, usually by direct measurement in a specialized facility. If radiometric measurements are made of all polarization states of the incident radiation, then the gain matrix can be inverted mathematically to correct for the off-diagonal contamination (Gasiewski and Kunkee, 1993). If only some of the states are measured, then the others must be estimated prior to the matrix inversion.

Receiver effects

A radiometer antenna collects incident radiation and converts it to a voltage on a transmission line. A radiometer receiver amplifies that voltage to a more measurable signal strength, typically with transistor amplifiers, and then measures the time average of the square of the voltage. This measurement is directly proportional to the power in the signal. The measured power contains one component due to the antenna temperature and another due to thermal emission by the antenna and receiver. The latter

component is called the receiver noise temperature (T_{rec}), and the sum of the two is referred to as the system noise temperature (T_{sys}), that is, $T_{sys} = T_a + T_{rec}$. The raw power measurement made by a radiometer is equal to T_{sys} multiplied by the receiver gain. T_{rec} originates primarily from three sources. Both the antenna and the transmission line connecting it to the amplifier are typically made of a highly conducting metal. However, because the conductivity is finite, there will be a small resistive loss in power experienced by the signal as its current flows along the metal. The metal will reradiate its own thermal emission in order to remain in local thermodynamic equilibrium. This reradiation contributes to T_{rec} . The amplifier also generates its own emission, due in part to resistive losses at its input stage as well as other noise generation mechanisms within the transistor. Other noise sources are also present in the receiver after its first amplifier, but they have a relatively small effect on T_{rec} because the T_a portion of the signal has been significantly amplified at that point.

Calibration of receiver effects consists of estimating the receiver gain and T_{rec} and then subtracting T_{rec} from T_{sys} to obtain T_a . T_{rec} can be estimated in different ways, depending on the radiometer hardware architecture. Many radiometer antennas are mechanically scanned to sweep their mainbeam across a T_b scene and generate an image. In this case, it is common to place a pair of calibration targets with known T_b values at opposite edges of the scan (Hollinger et al., 1990). Their measurement allows for the determination of both the receiver gain and T_{rec} . Some radiometers are not mechanically scanned and it is impractical to place calibration targets in front of the antenna, especially in spaceborne deployments. In these cases, receiver gain can be estimated by intermittently adding a known noise source to T_{sys} and measuring the resulting change in power. To estimate T_{rec} , a calibration switch is often used, which redirects the input to the receiver from the antenna to a calibration target with a known T_b (Ruf and Warnock, 2007). This approach can be less accurate and stable because the calibration switch itself is a source of power loss and thermal noise and because the resistive losses in the antenna and the portion of the transmission line between it and the receiver are now outside of the calibrated portion of the receiver.

Conclusion

Calibration of a microwave radiometer is the process by which the T_b arriving from a particular direction at a particular polarization is estimated from its raw measurements. The raw measurements can be explained by considering the action of the radiometer antenna on the incident radiation and then considering the action of the radiometer receiver on the signal generated by the antenna. The principle instrumental effects that are corrected for by the calibration process are the sensitivity of the antenna to incident radiation from undesirable directions and at undesirable polarizations, the addition

of thermal emission to the measurement by the receiver itself, and the magnitude of the signal amplification by the receiver.

Bibliography

- Gasiewski, A. B., and Kunkee, D., 1993. Calibration and applications of polarization correlating radiometers. *IEEE Transactions on Microwave Theory Techniques*, **29**(6), 1449.
- Hollinger, J. P., Pierce, J. L., and Poe, G. A., 1990. SSM/I instrument evaluation. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 781–790.
- Njoku, E. G., Christensen, E. J., and Cofield, R. T., 1980. The SeaSat scanning multichannel microwave radiometer (SMMR): antenna pattern correction – development and implementation. *IEEE Journal of Oceanic Engineering*, **OE-5**, 125.
- Ruf, C. S., and Warnock, A. M., 2007. GEOSAT follow on water vapor radiometer: performance with a shared active/passive antenna. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 970–977.

Cross-references

[Microwave Radiometers](#)
[Radiation, Polarization, and Coherence](#)

CALIBRATION, OPTICAL/INFRARED PASSIVE SENSORS

Carol Bruegge
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Definition

Calibration. A set of operations that establish, under specified conditions, the relationship between values indicated by a measuring instrument and the corresponding known values of a standard. For remote sensors, this typically implies the radiometric, spectral, and geometric characterization of an instrument as needed to understand the impact of the instrument's performance on the data or the derived data products.

Calibration factors are determined by comparison with a standard whose output is known in accepted physical units as part of the *Système International d'Unités* (abbreviated SI). Units are based on the metric-kilogram-second (mks) system and include the Kelvin, for temperature, and Watt, for power. Derived radiometric parameters are listed in [Table 1](#).

The calibration parameters derived for a sensor must be reported with the associated uncertainty and confidence level. The uncertainty analysis needs to establish both the precision and absolute uncertainty. The *precision* of the calibration is the consistency with repeated measurements. The *absolute calibration* is determined by the sensor's response to stable standards. These should be related to international or national standards through an unbroken chain of comparisons.

Calibration, Optical/Infrared Passive Sensors, Table 1
Common radiometric parameters

Entity	Symbol	Units
Spectral irradiance (radiant flux density at a surface)	E_λ	$\text{W m}^{-2} \mu\text{m}^{-1}$
Spectral radiance	L_λ	$\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$
Spectral reflectance (reflected flux/incident flux)	ρ_λ	Unitless

The *confidence level* is an interval about the result of a measurement within which the true value is expected to lie, as determined from an uncertainty analysis with a specified probability. A 3σ (sigma) confidence level implies that the stated uncertainty is achieved with 99 % probability.

Introduction

During the sensor development phase, the science and engineering teams first agree to set calibration requirements and a calibration approach. Absolute calibration requirements of 3–5 % (1σ) uncertainty are considered state of the art. Higher accuracy is typically specified for channel-relative or pixel-relative measurements. The accuracy is limited due to instrument-specific attributes, such as stray-light, out-of-field response, polarization sensitivity, scan-mirror sensitivity, linearity, signal-to-noise, temperature sensitivity, dark offset, and long-term stability. Radiometric uncertainty is increased when these effects alter the imagery and recorded signals in unpredictable ways.

Testing is conducted throughout the lifetime of the sensor. Characterizations of the filter and detector components provide early assurance that design requirements will be met. Preflight testing at the instrument level, before assembly onto the spacecraft, allows many parameters to be determined, which cannot be established on-orbit. This testing period is crucial to understanding the as-built performance. Following this, the program must commit to an on-orbit calibration plan. This allows response degradation, due to the browning of the optical elements or throughput change due to radiation damage, to be monitored. An adequate on-orbit calibration program will allow accurate radiance products to be reported even in the presence of sensor degradation. Both preflight and on-orbit calibrations are essential. An overview of the preflight testing for the Multi-angle Imaging SpectroRadiometer (MISR) is given by Bruegge et al. (2002).

Spectral response function

Most remote sensing systems measure incident light using several channels, where each channel is designed with a specified color sensitivity. That is, the instrument's per-channel output is designed to be a function of both the amount of incident light, as well as its wavelength. The color selection may be achieved by using spectral

filters within the instrument, or may be accomplished by use of a grating that disperses light according to the wavelength. In either case, the *spectral response function* (SRF) must be determined for each channel. This is a measure of the instrument's output as a function of the incident wavelength of light. The distribution only needs to be known in a relative sense, and is typically normalized to unity at the wavelength of peak response.

The SRF is measured using a test setup that can illuminate an instrument with monochromatic light, and where the wavelength of light can be varied during the test. The characterization should measure both the in-band response, near the region of peak sensitivity, and also the far-wing response. The hardware used is most commonly a monochromator. For instruments with very narrow spectral response functions, a tunable diode laser can be used, or Fourier Transform Interferometer (Strow et al., 2003). In either case, it is imperative that the system response be measured. It is insufficient to substitute the filter transmittance for this response, as all optical components, as well as the detector itself, contribute to the SRF.

It is common practice to summarize a sensor's spectral properties by tabulating the center wavelength and spectral width. This has been done by quoting the wavelength-of-peak response and full-width at half-maximum (FWHM). The latter is the wavelength at which the response falls to half of its peak value. These parameters can be misleading in cases where the SRF may be double peaked, asymmetric, or have a large out-of-band response. A better representation can be obtained from an equivalent square-band response analysis (Palmer, 1984). Here the sensor spectral response function is replaced with a function of equivalent area, but with an effective amplitude R_n and a throughput of 0 at wavelengths less than a computed minimum wavelength or greater than a computed maximum wavelength. Figure 1 gives an example of an SRF and square-band equivalent. The example is of Landsat-4, TM band 2 channel.

Radiometric calibration

Band-averaged radiances

Typically, raw data (digital numbers, or DN) from a sensor are converted to radiance units by knowledge of the gain coefficient. An example might be

$$DN - DN_0 = \bar{L}_\lambda \cdot G \quad (1)$$

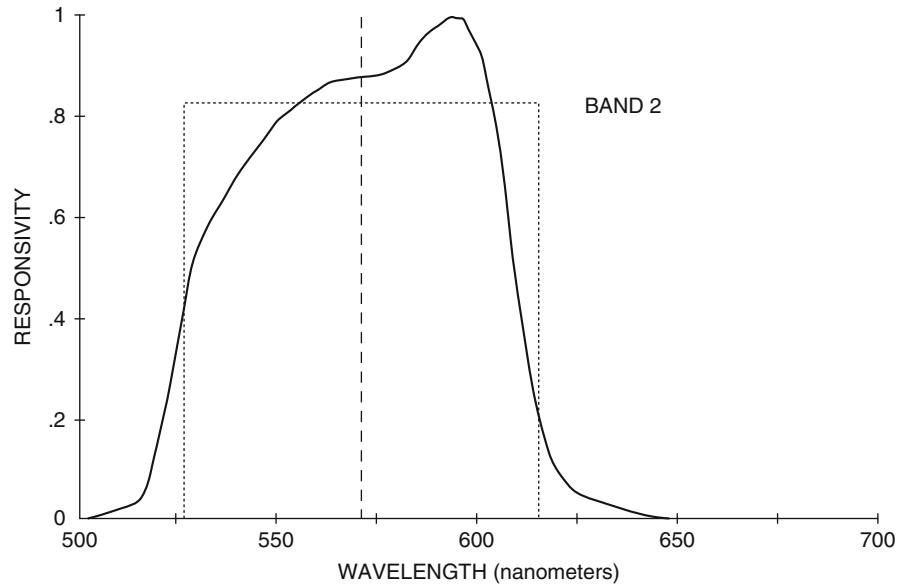
where

DN is the sensor output count

DN_0 is the dark-scene output count

$\bar{L}_\lambda$ is the band-weighted spectral radiance incident onto the sensor ($\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$)

and G is the radiometric response coefficient ($DN/[\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}]$)



Calibration, Optical/Infrared Passive Sensors, Figure 1 Spectral response function and moments-derived center wavelength and width for a typical detector.

A precise determination of the band-weighted radiance would require knowledge of both the spectral content of the incident light as well as the sensor spectral response function, R_λ . The band-weighted incident radiance is as follows:

$$\bar{L}_\lambda = \frac{\int L_\lambda R_\lambda \partial \lambda}{\int R_\lambda \partial \lambda} \quad (2)$$

For calibration, the gain, G , is computed with complete knowledge of the SRF and source spectral radiance. For radiance retrieval and science product generation, where output digital numbers are converted into radiances, the spectral content of the scene is not retrieved, rather only a band-weighted average.

Standards

The test equipment associated with a radiometric calibration requires an extended, spatially uniform, and spectrally smooth light source. A blackbody radiator is the most widely used source for infrared calibration. Its use in the visible, UV, and near-IR is limited. For these wavelengths, an integrating sphere or lamp and diffuse reflector fixture is more typical. The source must be larger than the geometric field of view of the sensor to be tested. This is to capture stray and diffracted light that will contribute to the output. Point sources of light are suitable for measuring a system's point spread function response, but are not suitable as radiometric targets.

If the source is an integrating sphere, its output needs to be calibrated by use of a transfer detector. The detector is, in turn, calibrated against a source standard. Standard

filament lamps are commercially available. The manufacturer typically seasons them by operating for 30 h. Lamps with output fluctuations are discarded. Suitable lamps are calibrated against a working standard that has a calibration traceable to national standards. These lamps can have accuracies approaching 1 % (2σ). Schneider and Goebel (1984) provide a review of standards.

On-orbit calibration

In-flight calibration is best accomplished using multiple technologies. Data from an on-board calibrator (OBC), for one, can provide the most frequent verification of sensor performance and stability. These systems can be used to make frequent checks on performance. Sampling dark current or the response to an attenuated view of the sun, for example, can be made once per orbit. As the OBC can itself degrade on-orbit, scene-viewing techniques are also required. Although considered the definitive validation of sensor performance, measurements are less frequent due to constraints associated with imagery collection. Scene studies can be as simple as observations of an un-instrumented desert site, or a highly accurate measurement of a site's in situ observations, and involving a field team or a network. Coastlines and contrasts edges can be used to confirm channel geolocation and co-registration. The SeaWiFS sensor makes use of lunar observations to track relative degradation changes (Eplee et al., 2001). Finally, cross-comparisons with sensors of similar footprints and bandpasses provide convincing error estimations. The comparison of radiances as determined from all techniques allows radiance uncertainty to be determined.

On-board calibrators

Examples of on-board calibrators can be found on EOS/Terra spacecraft sensors. The Multi-angle Imaging SpectroRadiometer (MISR) makes use of two deployable Spectralon diffuse targets. These are used to reflect sunlight into the earth-observing cameras. The panels have proven to be radiometrically stable on-orbit (Chrien et al., 2002). This is attributed to cleanliness and proper handling procedures that avoid exposure to contaminants (Stiegman et al., 1993). The panels are monitored by detector standards. Both radiation-resistant and high-quantum efficient detectors are utilized. The latter are intended to provide a measure of the incident light based upon physical principles, rather than transfer calibration using a standard source. In practice, detector stability has proven to be the driving criteria for detector selection.

The MODerate resolution Imaging SpectroRadiometer (MODIS), also on the Terra spacecraft, makes use of several on-board calibrator systems. These include a blackbody (BB) radiator, solar diffuser (SD), a solar-diffuse stability monitor, and the spectroradiometric calibration assembly (SRCA). The BB is the prime calibration source for the thermal bands located from 3.5 to 14.4 μm , while the SD provides a diffuse, solar-illuminated calibration source for the visible, near-infrared, and shortwave infrared bands ($0.4 \mu\text{m} \leq \lambda < 2.2 \mu\text{m}$). The SDSM tracks changes in the reflectance of the SD via reference to the sun so that potential instrument changes are not incorrectly attributed to changes in this calibration source. The SRCA is a very complex, multifunction calibration instrument that provides in-flight spectral, radiometric, and spatial calibration.

Unattended desert sites

The Sahara desert sites are considered stable with time. These sites have only small amounts of vegetation, are sparsely populated, and are typically found with clear-sky and low-aerosol conditions. Many instrument teams use these targets to monitor sensor degradation with time (Cosnefroy et al., 1996). Routine observations of sites such as Egypt_1 (26.10 East longitude, 27.12 North latitude) can be trended in order to determine the response degradation for a channel with time. Data are normalized by cosine of the solar zenith angle and the Earth–Sun distance. Observations are trended that are acquired at a fixed observation angle. This reduces error due to surface bidirectional reflectance factor (BRF) differences with view angle.

Vicarious calibration

Vicarious calibration (VC) is a process that is based upon in situ measurements acquired over a large, homogeneous desert site. With these data, the top-of-atmosphere spectral radiance can be computed and compared to those reported by the sensor as it simultaneously images the test site. Agreement constitutes

validation of the sensor's calibration. The in situ observations determine aerosol optical depths, surface spectral reflectance and BRF, and water vapor column amount. The top-of-atmosphere radiances are computed using a radiative transfer code such as MODTRAN. Vicarious calibrations can have uncertainties as small as 3 % (1σ). Key to the success of these measurements is the site selection. Western US desert sites, such as Railroad Valley, Nevada, are typically cloud-free and low in aerosols, minimizing errors in the radiative transfer calculation (Thome et al., 2003). It is for this reason that the vicarious calibration data is considered the radiometric standard for MISR, with the on-board calibrator adjusted to provide consistent calibrations (Bruegge et al., 2004).

Vicarious calibration requires visits by a field team to collect surface and atmospheric measurements at times necessarily coincident with a sensor over-flight. With the creation of an autonomous calibration facility, vicarious calibration data can be made available to the sensor community without the need for each research group to deploy its own field team. One example of an autonomous site is the LSpec (LED Spectrometer) automatic facility (Kerola et al., 2009). The facility is located at Frenchman Flat, within the Nevada Test Site. An array of eight LED spectrometers perform the autonomous function of recording surface reflectances at 5 min intervals, thereby permitting accurate and continual real-time scaling to a high-resolution characterization of the surface. Also resident at the LSpec site is a Cimel sun photometer, used to make atmospheric transmittance measurements. The Cimel is part of the Aerosol Robotic Network (AERONET; <http://aeronet.gsfc.nasa.gov/index.html>). Measurements made by the LSpec Cimel are used by AERONET to derive values of aerosol optical depths. From continuous in situ measurement of spectral reflectance of the playa surface, along with acquired aerosol optical depths and ozone optical depths (obtained from the Ozone Mapping Instrument OMI; <http://jwocky.nasa.gov>), the LSpec database provides all physical measurements needed to compute top-of-radiance with the same accuracy as traditional vicarious experiments. Data products are available to the public, and are available via a web-based interface at (<http://LSpec.Jpl.Nasa.Gov>).

Cross-calibration

The comparison of radiances from two or more sensors having similar passbands, acquired over a common target at near-coincident times can be a powerful validation exercise. In reality, the constraints on the measurements can be relaxed by making use of a radiance model for the site. That is, if the top-of-atmosphere spectral radiance can be determined, these data can be used to estimate the band-averaged radiance for a given sensor, at its SRF and for its time of observation. This is repeated for one or more sensors. The comparison of the first sensor

differenced to its model with the second sensor differenced to its model constitutes the cross-calibration (Thome et al., 2003).

Summary

There is an increasing demand for higher accuracy calibration. This is driven by the desire to have long-term data records that span multiple satellite programs. In an effort to establish best practices and minimize sensor-to-sensor biases, many international working groups and national standard laboratories are in collaboration. Information on the Committee on Earth Observation Satellites (CEOS) calibration/validation group can be obtained at <http://wgcv.ceos.org/>.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Bruegge, C. J., Chrien, N. L., Ando, R. R., Diner, D. J., Abdou, W. A., Helmlinger, M. C., Pilorz, S. H., and Thome, K. J., 2002. Early validation of the multi-angle imaging spectroradiometer (MISR) radiometric scale. *IEEE Transactions on Geoscience and Remote Sensing*, **40**(7), 1477–1492.
- Bruegge, C. J., Abdou, W. A., Diner, D. J., Gaitley, B. J., Helmlinger, M. C., Kahn, R. A., and Martonchik, J. V., 2004. Validating the MISR radiometric scale for the ocean aerosol science communities. In Morain, S. A., and Budge, A. M. (eds.), *Post-Launch Calibration of Satellite Sensors*. Leiden: A.A. Balkema, pp. 103–115.
- Chrien, N. L., Bruegge, C. J., and Ando, R. R., 2002. Multi-angle imaging spectroradiometer (MISR) on-board calibrator (OBC) in-flight performance studies. *IEEE Transactions on Geoscience and Remote Sensing*, **40**(7), 1493–1499.
- Cosnefroy, H., Briottet, X., and Leroy, M., 1996. Selection and characterization of Sahara and Arabian desert sites for the calibration of optical satellite sensors. *Remote Sensing of Environment*, **58**(1), 101–114.
- Eplee, R. E., Robinson, W. D., Bailey, S. W., Clark, D. K., Werdell, P. J., Wang, M., Barnes, R. A., and McClain, C. R., 2001. Calibration of SeaWiFS. II. Vicarious techniques. *Applied Optics*, **40**, 6701–6718.
- Kerola, D. X., Bruegge, C. J., Gross, H. N., and Helmlinger, M. C., 2009. Vicarious calibration of visible-near infrared earth remote sensors using LED Spectrometer (LSpec) facility measurements in conjunction with radiative transfer models. *IEEE Transactions on Geoscience and Remote Sensing*, **47**(4), 1244–125.
- Palmer, J. M., 1984. Effective bandwidths for LANDSAT-4 and LANDSAT-D' multispectral scanner and thematic mapper subsystems. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-22**(3), 336–338.
- Schneider, W. E., and Goebel, D. G., 1984. Standards for calibration of optical radiation measurement systems. *Laser Focus*, **20**(99), 82–96.
- Stiegman, A. E., Bruegge, C. J., and Springsteen, A. W., 1993. Ultraviolet stability and contamination analysis of Spectralon diffuse reflectance material. *Optical Engineering*, **32**(4), 799–804.
- Strow, L. L., Hannon, S. E., Weiler, M., Overoye, K., Gaiser, S. L., and Aumann, H. H., 2003. Prelaunch spectral calibration of the

atmospheric infrared sounder (AIRS). *IEEE Transactions on Geoscience and Remote Sensing*, **41**(2), 274–286.

Thome, K. J., Biggar, S. F., and Wisniewski, W., 2003. Cross comparison of EO-1 sensors and other Earth resources sensors to Landsat-7 ETM+ using Railroad Valley Playa. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6), 1180–1188.

Cross-references

[Calibration and Validation](#)

[Remote Sensing, Physics and Techniques](#)

CALIBRATION, SYNTHETIC APERTURE RADARS

Anthony Freeman

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

Imaging radar calibration

Definition

Synthetic aperture radar (SAR). A type of radar that forms high-resolution images of surfaces (planetary and terrestrial) using a technique known as aperture synthesis. *SAR calibration*. Encompasses all of the necessary steps taken to convert SAR image pixel values to fundamental units such as (normalized) radar cross section, measured in units of (m²/m²), or relative phase, measured in degrees. *SAR calibration performance*. An assessment of how well the calibrated SAR image pixel values correspond to the desired fundamental units.

Relative calibration performance (precision). An assessment of the relative errors between measurements made by a sensor of the same quantity separated in time or space. *Absolute calibration performance (accuracy)*. An assessment of the error in any single measurement made by a sensor as compared with an accepted, standard reference value.

Introduction

Synthetic aperture radar (SAR) was first conceived and demonstrated from aircraft in 1953 (Wiley, 1965). SARs are generally radars mounted on aircraft or satellite platforms and pointed sideways. The radar illuminates an area (referred to as its footprint) on the planet's surface down and off to the side of the platform track, and as the platform moves, this footprint traces out a swath that will eventually form the SAR image. SARs achieve high-resolution through aperture synthesis, a technique that combines radar returns collected from multiple (many) vantage points. This is done through signal processing, which generally occurs after the "raw" data has been sent down to the ground.

Early SARs were used to simply produce high-resolution images for military mapping purposes and later geologic or landform mapping. This was especially valuable in areas that were often cloud-covered and therefore unsuited to electro-optical sensor mapping. Interpretation and analysis of SAR images in this early period was not too different from photointerpretation of aerial photography and was therefore largely qualitative and highly subjective.

In the early 1980s, scientists using SAR data began to consider whether it could be used quantitatively, to allow geophysical quantities such as soil moisture, snow water equivalent, or surface topography (see also entry *Land Surface Topography*), for example, to be estimated directly from SAR measurements. Exciting results were being obtained from field experiments using well-calibrated, ground-based scatterometers with small footprints at around this time (Ulaby et al., 1982), and the expectation was that such local-scale results could be expanded to a global scale using SARs flown in space (Dubois et al., 1992). To achieve this, SAR data would have to be calibrated to standards similar to the ground-based scatterometers and to the same units, which had not happened up to that point. This challenge was vigorously tackled between 1987 and 1994 by scientists and engineers engaged in the field (Freeman, 1992).

As a result of all this effort, the first systematically calibrated SAR data became available to the science community in the late 1980s (van Zyl et al., 1992) from the NASA/JPL airborne SAR system (AIRSAR). This was followed in the early 1990s by calibrated data from the European Space Agency's ERS-1 spaceborne SAR (Attema and Francis, 1991), the Japanese Space Agency's JERS-1 SAR (Shimada, 1993), and NASA/JPL's SIR-C system in 1994 (Freeman et al., 1995). Since that time, calibrated SAR data has been the accepted norm, and systems such as the Canadian Space Agency's Radarsat or the Japanese Space Agency's PalSAR routinely deliver high-quality, calibrated SAR data to a broad science community.

External SAR calibration

During the 1980s, several groups around the world conducted experiments designed to calibrate SAR data collected over a given site. Examples include some of the field campaigns in support of the SIR-A (1981) and SIR-B (1984) missions, which carried a SAR in the payload bay of the space shuttle. Data from the SIR-A and SIR-B missions was largely uncalibrated, except for a few sites where calibration targets of known radar cross section were deployed and later used to estimate the radar cross section across the image (Way and Smith, 1991; Curlander and McDonough, 1991). This is referred to as external calibration. Calibration targets typically used in such experiments include trihedral corner reflectors, which resemble upturned pyramids in appearance and have a well-determined radar cross section, and

transponders, which can be tuned to have a range of radar cross sections and characteristics (Freeman et al., 1990).

Valuable as these field experiments were, it is of course impractical to deploy corner reflectors in every scene where one wants calibrated SAR data. Consider also that plans were under way in the 1980s to obtain high-resolution radar measurements of the surface of Venus and Titan by flying SARs on the Magellan and Cassini missions. What was needed was an approach that would allow systematic calibration of all data generated by a SAR sensor, without the need for calibration targets within the scene.

Internal SAR calibration

In the late 1980s, engineers began to adapt internal calibration methods developed for scatterometers (Ulaby et al., 1982), to allow the systematic conversion of all SAR image pixel values to fundamental units such as radar cross section. Internal calibration involved first constructing an error budget for the system that identifies all possible calibration error sources and their likely magnitudes. The next step is to design and build a radar system that is as stable as possible. All elements of the radar system that contribute to the calibration error budget and could therefore affect the final calibration performance are first characterized on the ground. Examples include the RF power measured in watts, which is transmitted by the radar when it illuminates the ground, and the gain of the receiver used to collect the reflected echoes from the area illuminated by the antenna footprint. SAR systems are somewhat unique in that the processing that turns the "raw" signal data collected by the SAR into high-resolution image data can also affect the calibration, so this therefore has to be factored into the error budget and characterized (Freeman and Curlander, 1989).

Preflight characterization of the individual elements of the SAR system is used to verify the system error budget. During flight operations, test signals are injected at reasonable intervals to check that the performance of the radar system is consistent with this preflight characterization. Examples of test signals are sinusoids or Gaussian noise sources covering a given RF frequency range, which are fed into the receiver. The RF power transmitted by the radar is also measured frequently. Some systems include capabilities to actively monitor the radar antenna performance (gain and pointing).

Putting all this together, systematic calibration of SAR data involves using the characterization of the end-to-end radar system from preflight and in-flight measurements, to arrive at a formula to convert the SAR image pixel values into fundamental units. The system error budget is then used to estimate the calibration performance for the final product. This calibration performance is verified using external calibration over sites containing calibration targets, or natural surfaces with sufficiently well-known radar reflectivity, such as the Amazon rain forest, or vegetation-free surfaces such as bare soil or open ocean.

Polarimetric SAR

One of the most challenging types of SAR data to calibrate comes from polarimetric radars, in which SAR images are collected simultaneously at several different polarizations. A typical polarimetric SAR system, such as NASA/JPL's AIRSAR, collects polarimetric SAR data by first transmitting a horizontally polarized radar waveform, then measuring the reflected energy from the surface in both horizontal (H) and vertical (V) polarizations. This produces radar images in what is termed HH and HV polarizations (H-transmit, H-receive and H-transmit, V-receive). The radar also transmits a vertically polarized waveform to produce images in VV and VH polarizations. H and V polarizations are orthogonal, and if one collects radar reflectivities in all four polarizations in this basis (i.e., HH, HV, VH, VV), then it is possible to synthesize what one would have seen with a radar configured with circular polarizations, for example (Zebker et al., 1987).

Polarimetric SAR data, besides allowing such polarization synthesis, are incredibly rich in information, and researchers have mined such data to retrieve estimates of soil moisture (Dubois et al., 1995), vegetation biomass (Dobson et al., 1992), and to differentiate between agricultural crops (Freeman et al., 1994). Polarimetric SAR data, when well-calibrated, can also be used to bound the underlying physics of the radar wave's interaction with the surface (Freeman and Durden, 1998) and to estimate structural characteristics (e.g., roughness scales) and electrical properties (dielectric constant or conductivity) of the surface under investigation (Oh et al., 1992).

Calibration of polarimetric SAR has its own particular set of challenges, one of which is that it is impossible to build a radar antenna that transmits or receives pure H or V polarizations. There is always some small contribution from the orthogonal polarization, termed cross talk, which has to be corrected or accounted for. van Zyl, (1990) was the first to derive a polarimetric calibration algorithm that used characteristics of naturally occurring surfaces to successfully correct for system cross talk. Another challenge is that the relative phase between polarizations (e.g., HH versus VV) contains significant information. Thus, approaches have been developed to calibrate relative phase between measurements (Sheen et al., 1989). Finally, interactions of the radar waveform with the ionosphere, particularly the effect of Faraday rotation (which changes the polarization of the radar waveform), are a significant source of calibration error at some wavelengths. Techniques to calibrate SAR data subject to Faraday rotation and at the same time other system effects appeared in the literature in 2004 (Freeman, 2004) and have since been validated using PalSAR data (Nicoll et al., 2007). Recently, interest has surfaced in a novel form of dual-polarized measurements known as compact or hybrid polarimetry, which carries less information than fully polarimetric SAR but has some advantages in terms of system design. Calibration of compact polarimetry systems presents some interesting challenges (Freeman et al., 2008).

Interferometric SAR

Interferometric SAR systems collect data from two different vantage points, and the relative phase between the two measurements is compared to estimate surface elevations or surface motion. Comparison of the relative phase between measurements allows observation of very subtle surface variations of the order of a fraction of the radar wavelength. Calibration of interferometric SAR then requires excellent phase calibration between measurements. These measurements may be made at the same time or over a time separation of years, in what is known as repeat-pass observations. This latter requirement introduces a long-term phase stability requirement into the system design.

Calibration of topography data derived from interferometric SAR data generated by the spaceborne radar topography mission (SRTM) was achieved by Rodriguez et al. (2006). Calibration of interferometric SAR data from different time periods was first demonstrated using ERS-1 data and has been demonstrated on a number of SAR missions since then. The largest calibration uncertainty in such repeat-pass observations stems from uncertainties in the relative location of the vantage points for each observation. The vector joining these two points is known as the interferometric baseline for that pair of observations. This baseline can vary from pass to pass, depending on the orbital characteristics of the platform and the degree to which the flight track can be controlled.

In recent years, techniques combining polarimetric and interferometric SAR have received a lot of attention (e.g., Dubois-Fernandez et al., 2005), with scientists publishing estimates of forest height derived from such data (termed PoLinSAR), using algorithms such as the random volume over ground model (Treuhaf et al., 1996). These techniques open up a new set of challenges for SAR calibration.

Conclusion

Since the late 1980s, calibrated SAR data has become the norm rather than the exception. Scientist using SAR data now expect it to be calibrated and to within known calibration uncertainties. Challenges exist in the calibration of SAR data from longer wavelength systems, subject to ionospheric effects, and when polarimetry and interferometry are combined. New science applications tend to push for better calibration performance, which also presents a challenge for the calibration engineer.

Acknowledgments

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Attema, E., and Francis, R., 1991. ERS-1 calibration and validation. *ESA Bulletin*, **65**, 80–86.

- Curlander, J. C., and McDonough, R. N., 1991. *Synthetic Aperture Radar Systems and Signal Processing*. New York: Wiley. Wiley Series in Remote Sensing.
- Dobson, M. C., et al., 1992. Dependence of radar backscatter on conifer forest biomass. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 3.
- Dubois, P. D., Evans, D., and van Zyl, J., 1992. Approach to derivation of SIR-C science requirements for calibration. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 6.
- Dubois, P. C., van Zyl, J., and Engman, T., 1995. Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 4.
- Dubois-Fernandez, P., Dupuis, X., and Garestier, F., 2005. PollnSAR calibration of deramp-on-receive system: example with the RAMSES airborne system. *Canadian Journal of Remote Sensing*, **31**, 1.
- Freeman, A., 1992. SAR calibration: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 6.
- Freeman, A., 2004. Calibration of linearly polarized polarimetric SAR data subject to Faraday rotation. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 8.
- Freeman, A., and Curlander, J. C., 1989. Radiometric correction and calibration of SAR images. *Photogrammetric Engineering and Remote Sensing*, **55**, 1295–1301.
- Freeman, A., and Durden, S., 1998. A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 3.
- Freeman, A., Shen, Y., and Werner, C. L., 1990. Polarimetric SAR calibration experiment using active radar calibrators. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 2.
- Freeman, A., Villaseñor, J., Klein, J. D., Hoozeboom, P., and Groot, J., 1994. On the use of multifrequency and polarimetric radar backscatter features for classification of agricultural crops. *International Journal of Remote Sensing*, **15**, 9.
- Freeman, A., et al., 1995. SIR-C data quality and calibration results. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 4.
- Freeman, A., Dubois, P. C., and Truong-Loi, M.-L., 2008. Compact polarimetry at longer wavelengths – calibration. In *Proceedings of EUSAR 2008*. Friedrichshafen.
- Nicoll, J., Meyer, F., and Jehle, M., 2007. Prediction and detection of Faraday rotation in ALOS PALSAR data. In *Proceedings of Geoscience and Remote Sensing Symposium*. Barcelona.
- Oh, Y., Sarabandi, K., and Ulaby, F. T., 1992. An empirical model and an inversion technique for radar scattering from bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 2.
- Rodríguez, E., Morris, C. S., and Belz, J. E., 2006. A global assessment of the SRTM performance. *Photogrammetric Engineering and Remote Sensing*, **72**, 3.
- Sheen, D. R., Freeman, A., and Kasischke, E. S., 1989. Phase calibration of polarimetric radar images. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-27**, 719–731.
- Shimada, M., 1993. J-ERS-1 SAR calibration and validation. In *Proceedings of CEOS SAR calibration workshop*. Noordwijk, The Netherlands.
- Treuhaft, R., Madsen, S., Moghaddam, M., and van Zyl, J., 1996. Vegetation characteristics and underlying topography from interferometric radar. *Radio Science*, **31**, 6.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1982. *Microwave Remote Sensing: Active and Passive*. London: Addison-Wesley. Remote Sensing.
- van Zyl, J. J., 1990. Calibration of polarimetric radar images using only image parameters and trihedral corner reflector responses. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 3.
- van Zyl, J. J., Carande, R., Lou, Y., Miller, T., and Wheeler, K., 1992. The NASA/JPL three-frequency polarimetric AIRSAR system. In *Proceedings of IGARSS '92*. Houston.
- Way, J., and Smith, E. A., 1991. The evolution of synthetic aperture radar systems and their progression to the EOS SAR. *IEEE Transactions on Geoscience and Remote Sensing*, **29**, 6.
- Wiley, C. A., 1965. Pulsed doppler radar methods and apparatus. U.S. Patent 3,196,436, Filed August 13 1954.
- Zebker, H. A., van Zyl, J. J., and Held, D. N., 1987. Imaging radar polarimetry from wave synthesis. *Journal of Geophysical Research*, **92**, 683–701.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)
[Radar, Synthetic Aperture](#)
[Radiation, Polarization, and Coherence](#)

CALIBRATION, SCATTEROMETERS

David Long

Department of Electrical and Computer Engineering,
 BYU Center for Remote Sensing, Brigham Young
 University, Provo, UT, USA

Definitions

Radar scatterometer. A calibrated radar designed to measure the radar backscatter cross section of a target, which is generally an area on the earth's surface.

Wind scatterometer. A scatterometer designed to measure the ocean's surface backscatter at multiple azimuth angles in order to estimate the near-surface vector wind. It is also used for ice melt/freeze, soil moisture, and vegetation remote sensing.

Normalized radar cross section. Area-normalized radar backscatter, which is the ratio of the incident and reflected radar signal power. Equivalent to the radar albedo. Often represented by σ^0 or sigma-o.

Geophysical model function (GMF). An empirical or analytic relationship between radar backscatter and geophysical quantity, that is, sigma-o is a function of the vector wind, polarization, and other parameters,

$$\sigma^0 = F(S, \chi, p, \dots)$$

where S is the neutral-stability wind speed (typically at 10 m height), χ is the relative angle between the radar illumination angle and the wind direction, p is the polarization (h or v), and $\dots$ represents other parameters such as the wave field, temperature, salinity, etc.

Introduction

A scatterometer is a carefully calibrated radar designed to measure the radar backscatter cross section (usually as the normalized radar backscatter coefficient) of a target, which is frequently part of a natural scene. Accurate calibration of the scatterometer-measured backscatter is critical to the utility of the scatterometer data, and much effort is expended to insure precise calibration of the radar. Scatterometer backscatter measurements are generally

used to infer geophysical properties about the target scene, for example, the near-surface wind speed, soil-moisture, or snow wetness. This is typically done with the aid of a geophysical model function (GMF) or functional relation that relates the parameter of interest to the radar backscatter. Thus, scatterometer calibration can also refer to the determination of this model function and the verification of the geophysical estimation from the scatterometer backscatter measurements. Note that radar parameters and the geometry used to collect the backscatter measurements must be accounted for in the GMF and that uncertainties in these parameters lead to errors in parameter estimation.

Scatterometer calibration fundamentals

Microwave scatterometers operate by transmitting a microwave signal toward the target and measuring the reflected or backscattered power; the radar cross section is computed from the backscattered power measurement. Due to thermal noise in the radar receiver and Rayleigh fading, the signal power measurement is corrupted by noise. A separate measurement of the receive-only noise power is made and subtracted from the signal-plus-noise measurement to estimate the signal power measurement. The signal-to-noise ratio (SNR) may vary from very low (less than -20 dB) to high (20 dB or more).

For areal (scene) targets, the backscattered signal power is related to the normalized radar cross section (σ^0) via the radar equation (Ulaby et al., 1981), which can be written as

$$\sigma^0 = P_s X \quad (1)$$

where P_s is the received backscatter power, and the backscatter conversion factor X (in simplified form) is

$$X = g P_t G^2 \lambda^2 A_c / (4\pi)^3 R^4 L \quad (2)$$

where processor g is a calibration correction, P_t is the transmitted power, G is antenna gain, λ is the wavelength of the transmitted microwave signal, A_c is the illuminated footprint area, R is the slant range to the surface, and L represents system losses, including processing losses. A key goal in scatterometer calibration is ensuring accurate computation of the backscatter cross section by selecting appropriate values for g and X .

More accurate expressions for X include an integral of the radar equation parameters over the surface, for example,

$$X = \frac{P_t \lambda}{(4\pi)^3 L} \iint g(x, y) G(x, y) / R^4(x, y) dx dy \quad (3)$$

where the integral is defined over the non-negligible antenna gain and x and y are orthogonal surface coordinates. (Other forms may be more convenient for a given application such as expressing the integral in terms of antenna angles.)

The parameters in X can be split into two categories: those that depend on the observation geometry (G , A_c , R , and possible contributions to g and L) and those that depend on the radar design and RF system gain (P_t , λ , and possible contributions to g and L). Thus, scatterometer backscatter calibration requires both geometric and radar parameter procedures. The parameters g and L include factors such as signal spill-over and clipping loss from the radar range gates, signal processing filters, the processing scheme employed, etc. These parameters are highly instrument-specific; techniques for determining them must be adapted for each sensor.

Backscatter calibration

The scatterometer backscattered signal from the target scene consists of distance-attenuated, time-delayed, frequency-shifted copies of the transmitted signal. The time delay arises due to finite speed of light and the distance (sometimes referred to as the slant range) between the radar and the scene. Variations in the time delay due to spacecraft pointing, topography, and orbit can result in signal power loss by the range gate. The relative motion of the radar with respect to the scene introduces Doppler frequency shift, and variations in Doppler can introduce power variations due to nonideal filters. These must be accounted for in computing X . Uncertainties or errors in computing X affect both the mean and variance of the σ^0 measurement. Scatterometer σ^0 measurement calibration includes insuring that both the mean and variance in the estimated σ^0 error in the measurement are minimized. This is complicated by measurement errors due to radiometric noise and Rayleigh fading, which results from time-fluctuation of the self-interference in the echo signal (Ulaby et al., 1981, 1990).

A model for the measured backscatter σ_m^0 observed by the scatterometer is

$$\sigma_m^0 = \sigma^0 (1 + K_p v) \quad (4)$$

where σ^0 is the true or expected backscatter of the surface, K_p is the normalized standard deviation (sometimes referred to as the radiometric accuracy) of the measurement error, and v is typically a normally distributed Gaussian random variable. The multiplicative noise model in Eq. 4 arises, in part, as a result of speckle noise, an inherent limitation in coherent radars. The measurement variability K_p can be expressed as (see also Yoho and Long, 2003, 2004)

$$K_p^2 = K_{pc}^2 + K_{pr}^2 + K_{pm}^2 \quad (5)$$

where K_{pc} is the “communications” variability due to radiometric noise, K_{pr} is the calibration uncertainty and variability, and K_{pm} is the modeling variability. A key goal in scatterometer calibration is to minimize K_{pr} and K_{pm} . K_{pc} can be written as a quadratic function of the true signal-to-noise ratio ($\text{SNR} = P_s/P_n$ where $P_s = X_t \sigma^0$ is

the signal power, X_t the true value of X , and P_n is the noise power),

$$K_p^2 = \alpha + \beta/\text{SNR} + \gamma/\text{SNR}^2 \quad (6)$$

where α , β , and γ depend on the radar hardware and the signal processing used (e.g., Fischer, 1972; Long and Mendel, 1991; Naderi et al., 1991; Yoho and Long, 2003, 2004).

Accurate ground-based calibration of the RF hardware is required to compute amplifier and antenna gains, filter responses, timing, and other parameters affecting X and the K_p parameters α , β , and γ . Scatterometers employ carefully designed and calibrated signal feedback paths that attenuate the transmit signal and feed it into the receiver. This permits monitoring and frequent internal calibration of the end-to-end RF system gain. Separate calibration and characterization of the antenna gain pattern must be performed prelaunch.

Once on-orbit, the total system gain including the antennas can be verified by using ground calibration stations. Receive-only, receive/transmit, and transponder ground stations have been used. Transponder stations receive the signal, amplify it by a fixed amount, (optionally) frequency-shift it, and retransmit it toward the orbiting instrument. This provides a signal that stands out against the ordinary backscatter in the measurement data, which enables the system signal timing and ground station received power to be checked against predicted values. As the sensor flies overhead, data from a fixed ground station provides a “slice” through the antenna pattern and measurement swath. Multiple stations enable multiple simultaneous slices to verify and calibrate antenna gain along fan-beam antenna patterns. Typically, ground stations use broad-beam horn antennas that are steered toward the scatterometer as it flies overhead in order to minimize possible signal gain variations in the ground station antennas. With care, absolute system calibration can be accomplished. However, while ground stations are effective in verifying calibration and providing on-orbit adjustments to antenna patterns, achieving the desired high-precision absolute system calibration (better than 0.1 dB) can be difficult since the calibration precision of the ground station is difficult to achieve and maintain.

A widely used alternate approach to system calibration is the use of natural targets. These regions have been used as calibration transfer standards to ensure accurate cross-calibration of multiple instruments. Distributed area targets that have a spatially uniform, temporally stable backscatter response are desired. Rain forests have been widely used for scatterometer calibration due to their wide areal extent and low seasonal variation, but other regions have also been used, including firm at the summits of Greenland and Antarctica and (during certain seasons) deserts and Asiatic steppes. The absolute backscatter of these regions is not known precisely, so natural scenes are generally used for relative calibration of the

backscatter versus observation geometry (such as azimuth angle or intra-footprint) over the scatterometer measurement swath. Rain forests have proven to be the most effective in intersensor calibration and as transfer standards in cross-calibrating difference sensors. Such areas are useful for long-term monitoring of sensor drift.

In the case of dense vegetation, the radar cross section is primarily dependent on the canopy density and varies nearly linearly (in dB) with measurement incidence angle. Carefully selecting homogenous areas within the rain forests that have minimal seasonal variations provides constant regions with the desired calibration characteristics. While the Amazon and Congo rainforests have been widely used, the rainforests of Borneo have proven less suitable for scatterometer calibration due to the mountainous terrain. Seasonal and daily variations in canopy moisture content, particularly daily cycles of dew, introduce some variability in the forest backscatter. For a fixed time of day, the areal average backscatter variation is estimated to be ± 0.15 dB at both Ku- and C-band (Long and Skouson, 1995; Long, 1998). Excluding measurements made in areas with recent (within 24–48 h) rain reduces the variability. Considerations for incidence and azimuth angle variations are essential, particularly in deserts and snow-covered areas, due to the presence of dunes that exhibit strong azimuthally dependent scattering characteristics.

We note that because of its narrow footprint, pencil-beam scatterometers such as SeaWinds on QuikSCAT receive-only measurements can function as radiometer observations useful for detecting rain. Calibrating QuikSCAT radiometric measurements requires converting the receive-only power measurements into brightness temperature values. Prelaunch RF calibration provides key receive system calibration parameters. Post-launch, comparisons of collocated brightness temperatures measured over extended area targets by conventional radiometers are used in order to complete the calibration process (Jones et al., 2000).

GMF calibration

Scatterometers are designed to accurately measure σ^0 in order to infer geophysical properties from the σ^0 measurements with the aid of a GMF that relates σ^0 with its observation geometry and the geophysical property of interest, for example, the near-surface ocean wind. An important component of scatterometer calibration is thus to ensure the accuracy of this indirect estimation process by “calibrating” the GMF. In this context, “calibration” refers to validating the accuracy of the estimated geophysical property, for example, the surface wind, by comparison to independent measurements, for example, from buoys. The GMF or backscatter calibration is adjusted to minimize the error (and variance) of the estimated geophysical parameter.

GMF calibration is conducted with the aid of campaigns to collect surface parameter data that is

compared to spatially and temporally collocated backscatter measurements. An extensive literature exists on this topic. Ideally, the surface measurements span the expected range of the parameter of interest and parameters that may affect it. For example, for wind calibration, buoy and/or ship measurements are collected at multiple locations with varying sea-surface temperature, wave conditions, and locations within the scatterometer swath. The comparison data is segmented in various ways in order to evaluate potential sensitivity to unmodeled factors, for example, sea-surface temperature or significant wave height. Near-surface wind measurements are controlled or adjusted to similar heights, fetch, and atmospheric stability conditions using bulk corrections. Assumptions of ergodicity are required comparing buoy point measurements to areal average scatterometer measurements. Rain and land proximity adversely affect the accuracy of scatterometer-derived wind estimates and so only open ocean and no-rain surface data are considered in wind GMF calibration. The size of the scatterometer footprint alters the distribution of the backscatter, particularly in low-wind conditions that can result in very low backscatter. At the other extreme, high winds can saturate the GMF, leading to reduced wind estimate accuracy.

An alternate GMF calibration approach relies on comparison of winds generated from numerical weather prediction (NWP) models and collocated scatterometer backscatter values. While all NWP models have accuracy limitations, the underlying assumption of the alternate approach is that if enough collocated backscatter measurements and NWP winds are collected, the NWP errors "average out," enabling estimation of an empirical GMF by binning the backscatter data according to the NWP output value. Though not without limitation, this technique has proven highly successful for scatterometer GMF calibration. GMFs, to support simultaneous wind and rain estimation from scatterometer backscatter measurements, rely on both NWP winds and collocated measurements from other sensors, such as radiometers or rain radars, to derive rain-effect and correction models.

Conclusion

Scatterometry has proven to be a very useful tool for operational wind measure and monitoring and studying the earth. Scatterometer data complements the data from other sensors in long-term studies. An essential element in the success of scatterometry is the precise calibration of the data, which accounts for the detail and attention scatterometer calibration is given.

Bibliography

Anderson, C., Bonekamp, H., Wilson, J., Figa, J., de Smet, J., and Duff, C., 2007. Calibration and validation of ASCAT backscatter. In *Proceedings of the Joint 2007 EUMETSAT Meteorological Satellite Conference and the 15th Satellite Meteorology & Oceanography Conference of the American Meteorological Society*, Amsterdam.

- Fischer, R. E., 1972. Standard deviation of scatterometer measurements from space. *IEEE Transactions on Geoscience Electronics*, **GE-10**(2), 106–113.
- Grantham, W. L., Bracalente, E. M., Britt, C. L., Wentz, F. J., Jr., Jones, W. L., Jr., and Schroeder, L. C., 1982. Performance evaluation of an operational spaceborne scatterometer. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-20**(3), 250–254.
- Jones, W. L., Merhershahi, R., Zec, J., and Long, D. G., 2000. SeaWinds on QuikSCAT radiometric measurements and calibration. In *Proceedings of the International Geoscience and Remote Sensing Symposium*, July 24–28, 2000, Honolulu, HI, pp. 1027–1029.
- Long, D. G., 1998. Comparison of TRMM and NSCAT observations of surface backscatter over the Amazon rainforest. In *Proceedings of the International Geoscience and Remote Sensing Symposium*, July 6–10, 1998, Seattle, WA, pp. 1879–1881.
- Long, D. G., and Mendel, J. M., 1991. Identifiability in wind estimation from wind scatterometer measurements. *IEEE Transactions on Geoscience and Remote Sensing*, **29**(2), 268–276.
- Long, D. G., and Skouson, G. B., 1995. Calibration of spaceborne scatterometers using tropical rainforests. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(2), 413–424.
- Naderi, F., Freilich, M. H., and Long, D. G., 1991. Spaceborne radar measurement of wind velocity over the ocean—an overview of the NSCAT scatterometer system. *Proceedings of the IEEE*, **79**(6), 850–866.
- Offiler, D., 1994. The calibration of ERS-1 satellite scatterometer winds. *Journal of Atmospheric and Oceanic Technology*, **11**(4), 1002–1017.
- Quilfen, Y., and Bentamy, Y., 1994. Calibration/validation of ERS-1 scatterometer precision products. In *Proceedings of the Geoscience and Remote Sensing Symposium*, August 8–12, 1994, Pasadena, CA, pp. 945–947.
- Spencer, M. W., Wu, C., and Long, D. G., 2000. Improved resolution backscatter measurements with the SeaWinds pencil-beam scatterometer. *IEEE Transactions on Geoscience and Remote Sensing*, **38**(1), 89–104.
- Tsai, W. Y., Graf, J. E., Winn, C., Huddleston, J. N., Dunbar, S., Freilich, M. H., Wentz, F. J., Long, G. J., and Jones, W. L., 1999. Postlaunch sensor verification and calibration of the NASA scatterometer. *IEEE Transactions on Geoscience and Remote Sensing*, **37**(3), 1517–1542.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing – Active and Passive*. Reading, MA: Addison-Wesley, Vols. 1 and 2.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1990. *Microwave Remote Sensing – Active and Passive*. Reading, MA: Addison-Wesley, Vol. 3.
- Verspeek, J., Stoffelen, A., Portabella, M., Verhoef, A., and Vogelzang, J., 2007. ASCAT scatterometer Cal/Val. In *Proceedings of the Joint 2007 EUMETSAT Meteorological Satellite Conference and the 15th Satellite Meteorology & Oceanography Conference of the American Meteorological Society*, Amsterdam.
- Yoho, P. K., and Long, D. G., 2001. Model-based ground station calibration for SeaWinds on QuikSCAT. In Barnes, W. L. (eds.), *Earth observing systems IV, Proceedings of SPIE*, Vol. 4483, 29 July–3 August, San Diego, CA.
- Yoho, P. K., and Long, D. G., 2003. An improved simulation model for spaceborne scatterometer measurements. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(11), 2692–2695.
- Yoho, P. K., and Long, D. G., 2004. Correlation and covariance of satellite scatterometer measurements. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(6), 1176–1187.
- Zec, J., Jones, W. L., and Long, D. G., 1999. NSCAT normalized radar backscattering coefficient biases using homogenous

land targets. *Journal of Geophysical Research*, **104**(C5), 11557–11568.

Zec, J., Jones, W. L., and Long, D. G., 2000. SeaWinds beam and slice balance using data over Amazonian rainforest. In *Proceedings of the International Geoscience and Remote Sensing Symposium*, July 24–28, 2000, Honolulu, HI, pp. 2215–2217.

Cross-references

[Radar, Scatterometers](#)

CLIMATE DATA RECORDS

Eric F. Wood

Department of Civil and Environmental Engineering,
Princeton University, Princeton, NJ, USA

Definition

Following NRC (2004), a climate data record (CDR) is defined as *a time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change*. For satellite-based CDR, these can be further defined as *fundamental CDRs (FCDRs)*, which are calibrated and quality-controlled sensor data that have been improved over time, and *thematic CDRs (TCDRs)*, which are geophysical variables derived from the FCDRs, such as [sea surface temperature](#) and cloud fraction.

Applying the nomenclature that a satellite record meets the standard of a CDR implies that the products were generated and maintained with attention to data stewardship and management, with developed access and dissemination policies. CDR often requires updating and reprocessing to maintain consistency and quality, which implies that the underlying datasets and information used in their creation, such as metadata, are preserved indefinitely in formats that promote easy access. Products deemed a CDR requires commitment for long-term data storage, access, and management.

Key elements of a climate data record activity

CDR organizational elements

1. An advisory and review team to provide input to development of a CDR (FCDR) from satellite data, to review the theoretical basis for the CDR (FCDR) algorithm, and to review the generated CDR (FCDR) products.
2. An advisory and review team for a fundamental CDR should have specific expertise for that product. For thematic CDRs (TCDRs), advisory and review teams should reflect broad disciplinary theme areas.

CDR generation elements

1. FCDRs must be generated with the highest possible accuracy and stability.
2. Sensors must be thoroughly characterized before and after launch, and their performance should be continuously monitored throughout their lifetime.

3. Sensors should be thoroughly calibrated, including nominal calibration of sensors in orbit, vicarious calibration with in situ data, and satellite-to-satellite cross-calibration.
4. TCDRs should be selected based on well-defined criteria established by the advisory and review team.
5. Validated TCDRs must have well-defined levels of uncertainty.
6. An ongoing program of correlative in situ measurements is required to validate TCDRs.

Sustaining CDR elements

1. Resources should be made available for reprocessing the CDRs as new information and improved algorithms are available, while also maintaining the forward processing of data in near real time.
2. Provisions should be included to receive feedback from the scientific community.
3. A long-term commitment of resources should be made to the generation and archival of CDRs and associated documentation and metadata.

Acknowledgments

The material in this section was drawn from NRC (2004).

Bibliography

NRC, 2004. *Climate Data Records from Environmental Satellites: Interim Report (2004)*. Washington, DC: Committee on Climate Data Records from NOAA Operational Satellites/National Research Council/National Academies Press. 150 p, 7 x 10.

Cross-references

[Sea Surface Temperature](#)

CLIMATE MONITORING AND PREDICTION

Mathew R. P. Sapiano

University of Maryland - College Park, College Park,
MD, USA

Definition

Climate. Time average of atmospheric properties. The objective of the temporal averaging is to establish a background state of behavior that is the most likely state of the atmosphere at any given time. Weather variations are considered to be superimposed on the background climate, while at any given location the climate is viewed as changing through the year; thus, a location has a different climate state for each month of the year. Spatial averaging is generally used as well, but if sufficiently detailed observations are available, microclimates of any desired scale can be defined. Climate variability is the year-to-year change in climate at a given location and time of year; if temporal and/or spatial coherence can be identified in

these variations, coherent modes of climate variability such as the El Niño/Southern Oscillation (ENSO) or the North Atlantic Oscillation (NAO) may be described.

Climate monitoring. The practice of describing the current state of the climate system and the status of coherent aspects of climate variability, generally through the use of temporally averaged atmospheric observations.

Climate prediction. Prediction of the future state of the climate system on seasonal to interannual timescales, often distinguished from climate change projections on longer timescales. An overlap exists between these concepts on timescales of several years to several decades.

Introduction

Mark Twain provided the most succinct definition of climate by defining its relationship with weather: “Climate is what we expect, Weather is what we get.” Contrasting weather with climate is perhaps the best approach to defining climate. The weather we observe is deterministic and dynamical; there is a single realization of the weather that is/was actually observed, albeit subject to some measurement error. Climate, on the other hand, is inherently statistical and is often thought of as the *mean state of the weather*, where this mean can be either in time or space (and commonly both). Additionally, weather can be thought of as being deterministic and climate is probabilistic, in that it is the most likely outcome out of some range of possible values (note that the mean usually satisfies the criteria of being the most likely outcome). For example, we might know that it rained on some given day (although there is, of course, measurement error on such an estimate), but the climate estimate of whether it rained on any given day might be a probability of rain rather than an outright answer. For climate monitoring and prediction, the mean underlying state of the Earth system at some given timescale (say the seasonal scale) is of prime interest.

In practical terms, weather monitoring, climate monitoring, and climate change monitoring are often differentiated by the timescale of the signal of interest which is important as a control of the allowable noise/error levels in the data. Weather scales include high-frequency variations and are typically defined as being between zero and 10 days, with longer timescales being more typically associated with climate signals. Climate change signals are generally small and low frequency and are defined as occurring over multiple decades or centuries. Climate monitoring includes scales longer than a few days but still within the period of a few months or years and generally involves the averaging of multiple weather events. Examples of this might be an enhanced hurricane season, a prolonged drought, or even a more expansive hole in the ozone layer.

Despite the many engineering and technical challenges associated with remote sensing datasets, they play a vital role in climate monitoring. Before remote sensing,

observations were limited to single point locations which were generally made in well-populated areas. Remote sensing through radar and satellite observations provides far increased spatial coverage as well as enhanced temporal sampling. This has allowed exploration of the climate in areas of little or no knowledge as well as an understanding of the processes acting at sub-daily timescales. The latency (time until availability) of remotely sensed is also ideal for climate monitoring. Traditionally, climate monitoring products were compiled months after the fact, but remotely sensed data is often available in real time so that global monitoring datasets can be available in mere hours. The construction of climate monitoring datasets from radar networks provides a significant technical challenge and few exist (a notable exception is the Stage IV US gauge-corrected radar precipitation estimates; Lin and Mitchell, 2005). Satellite observations, however, have proven crucial in both climate monitoring and climate change studies and have filled significant gaps in our records most notably, the global oceans.

For the purposes of this entry, we characterize remotely sensed observations in four groups in order from the most direct to the least direct:

1. Direct measurements of climate parameters, specifically the radiation budget at the top of atmosphere
2. Inferred measurements of the current state of important climate characteristics, such as clouds, SST, vegetation, and moisture
3. Contributions to the definition of the initial state of the climate system for prediction or reanalysis models
4. Verification data to for the assessment of climate models and other products and their ability to predict characteristics of the climate system

These groupings are by no means mutually exclusive, with the third and fourth making considerable use of direct and inferred measurements of climate parameters from the first and second group.

We will give examples of the first three groups in the following sections; the fourth is neglected since this is not directly relevant to climate monitoring and prediction, although it is a necessary part of product development. The vast majority of climate monitoring is achieved using inferred observations which fall under the second group, so this section is the largest and has several examples. Remotely sensed data is used in a wide range of different monitoring and prediction capacities; thus, a full survey of all possible activities is impossible. Likewise, each monitoring or prediction product utilizes different techniques, sensors, and algorithms which are too varied to allow for simple classification.

Measurements of the radiation budget

Spaceborne remotely sensed observations directly measure emitted or reflected radiation at the top of the atmosphere (TOA). Thus, the most direct measurements available from satellites are of the radiation budget which

includes the surface albedo, reflected shortwave radiation from the Sun, and emitted long-wave radiation (usually from the Earth's surface or clouds). While the albedo is fixed for a given surface type, the latter two are modulated by the Earth system. The TOA radiation budget is known to be critically important for climate: Global climate change is a reaction to changes in the energy budget due to increased greenhouse gas concentrations in the atmosphere. Despite the importance of the radiation budget to the climate system, it is still not fully understood how short-term perturbations affect the climate, and, as such, radiation balance estimates are not directly used in climate monitoring. However, estimates of the radiation budget are used to assess the performance of climate models, and so these observations are directly used in that context.

Example 1: total solar irradiance

Estimates of incoming solar radiation were traditionally made through the proxy of sunspot counts (see [Radiation, Solar and Lunar](#)). Sunspots are magnetic disturbances on the Sun's surface that have lower temperatures than their surroundings and lead to slight variations in solar activity with a regular 11 year cycle. Total solar irradiance (TSI) is a more direct measurement of solar activity and is defined as the wavelength-integrated solar radiation received by a surface at an average distance from the Sun at the top of Earth's atmosphere. Measurements of TSI are not possible from Earth's surface due to the effects of the atmosphere on radiation and so measurement of TSI was only made possible with the advent of satellite observations. The historical record of TSI is comprised of several different radiometers aboard different missions (see Fröhlich and Lean, 2004 or Fröhlich, 2007, for details). Multiple combinations of these datasets are available, each with different calibrations (Fröhlich, 2007). The most recent sensor to be launched is the Total Irradiance Monitor aboard the Glory satellite (Mishchenko et al., 2007) which will continue the record of TSI measurements into the future.

TSI is an example of a parameter which is primarily monitored for climate change detection and attribution studies. The "standard" value of TSI is approximately $1,366 \text{ Wm}^{-2}$ although absolute calibration of sensors is extremely challenging and disagreement exists over the mean value. Over the course of the solar cycle, small variations occur of the order of 2 Wm^{-2} (de Toma et al., 2004; Rottman, 2006). Variations in TSI are known to affect climate, and there has been much debate as to whether trends in TSI (or other measures of solar activity such as spectral changes or changes in cosmic rays) could explain recent increases in global temperature. The IPCC AR4 Report (Forster et al., 2007) showed that the contribution of TSI is small and that it is not the main driver of observed changes, although TSI does account for some of the increase.

While TSI affects almost all climate parameters, its effects are greatly obscured by its indirect nature: Solar radiation is the primary driver of the Earth system but it

is modulated by that system. Daily data are routinely available from the web, such as data from the Solar Radiation and Climate Experiment (SoRCE), and can be used in conjunction with other data to understand major solar events such as solar flare activity (e.g., Woods et al., 2004) which impact the Earth. Despite the direct link with the Earth system, it is unclear as to whether TSI variations are large enough to affect the Earth system in a medium-range sense and how to interpret and use these observations for climate monitoring.

Measurements of climate parameters

Estimates of Earth system parameters derived from satellites are based on inferences made from measurements of TOA radiation in different channels. The exact nature of these inferred measurements is different for each parameter, mission, sensor, and orbit type. Many parameters require the use of several channels to derive a more accurate estimate, and multiple satellites are often used. For instance, precipitation estimates usually use a mixture of geostationary infrared and polar-orbiting passive microwave estimates to achieve the best global estimate. Multisatellite estimates are commonly used in climate monitoring since these are often superior for surface parameters. The following examples are arranged in order from TOA to the surface and are not intended to be exhaustive, but it is hoped that they are sufficient to provide the interested reader with a foundation for further reading.

Example 2: ozone

In the latter part of the twentieth century, before the concept of global climate change had entered the public psyche, there was great public alarm over the discovery of a decline in [stratospheric ozone](#) and the existence of a hole above Antarctica in the thin ozone layer (see [Stratospheric Ozone](#)). Ozone is a gas residing primarily in the stable part of the atmosphere above the turbulent troposphere called the stratosphere. Stratospheric ozone occurs naturally at 15–35 km above Earth's surface where ultraviolet (UV) radiation from the Sun fuels a constant cycle of creation and destruction which keeps ozone levels in equilibrium. The ozone layer therefore absorbs nearly all UV radiation which can be harmful to human life, leading to a range of issues including forms of skin cancer. It was discovered in the 1970s that some man-made substances including chlorofluorocarbons (CFCs) were reaching the stable stratosphere and destroying ozone faster than it could be replenished. The Montreal Protocol was established in 1987 (and amended several times since) with the aim of phasing out the use of certain man-made chemicals known to destroy ozone. The treaty has been widely hailed as a success, and global stratospheric ozone levels have begun to recover, although the seasonal Antarctic ozone hole continues to exist and is not expected to recover to pre-1980 levels until midway through the twenty-first century (WMO, 2007).

This ongoing environmental issue is an example of how climate monitoring can have a major impact on society. In situ observations were made throughout the twentieth century, but the emergence of satellite observations with near-global coverage gave the data required to force policy makers into action. The Solar Backscatter Ultraviolet instrument (SBUV/2) aboard the NOAA series of satellites measures scattered UV rays to produce estimates of the total ozone as well as the vertical profile (Bhartia et al., 1996). The European Space Agency (ESA) Global Ozone Monitoring Experiment (GOME) instrument also provides total column ozone and the ozone vertical profile (Burrows et al., 1999) based on backscattered UV radiation.

The United Nations Environment Programme (UNEP) Ozone Secretariat is the primary international body for ozone monitoring. They have produced several assessment documents including the *Scientific Assessment of Ozone Depletion* series of reports which form the basis of global ozone monitoring. Other agencies such as the NOAA Climate Prediction Center (CPC) and other country-level meteorological agencies monitor the current state of stratospheric ozone.

Ozone is also found in much lower concentrations at low levels from anthropogenic sources. This so-called tropospheric ozone is regarded as a pollutant and can lead to serious, direct human health repercussions such as respiratory problems. The mechanisms leading to the formation of tropospheric ozone are different to those of stratospheric ozone, but similar techniques can be employed to retrieve estimates from remotely sensed information. While these estimates are not used in a climate monitoring sense, they are used in air quality forecasting.

Example 3: precipitation

Accurate measurements of precipitation are challenging because of the high variability of rainfall in both time and space (see [Rainfall](#)). Significant individual rainfall events can be as short as a few minutes in the case of highly localized convection or last several hours or days and be spread over a large area, as is the case for large-scale stratiform events. Differences in the relative skill of these precipitation estimates are frequently related to the ability to continuously sample precipitation so as not to miss short or small events. For that reason, gauges are usually considered the best source of precipitation data since they provide a time-integrated observation and have few calibration errors. However, gauges can be biased by wind-induced under-catch or other external factors, and coverage is often limited to populated areas where observers can access them. Ground radar estimates give good sampling over a limited area, but calibration issues preclude their combination to form a truly global dataset (the Stage IV radar dataset is a notable exception for the United States only). Satellite estimates of precipitation have larger sampling errors than either gauges or ground radar estimates but give the only estimates over the ocean. Satellite precipitation estimates are based on either

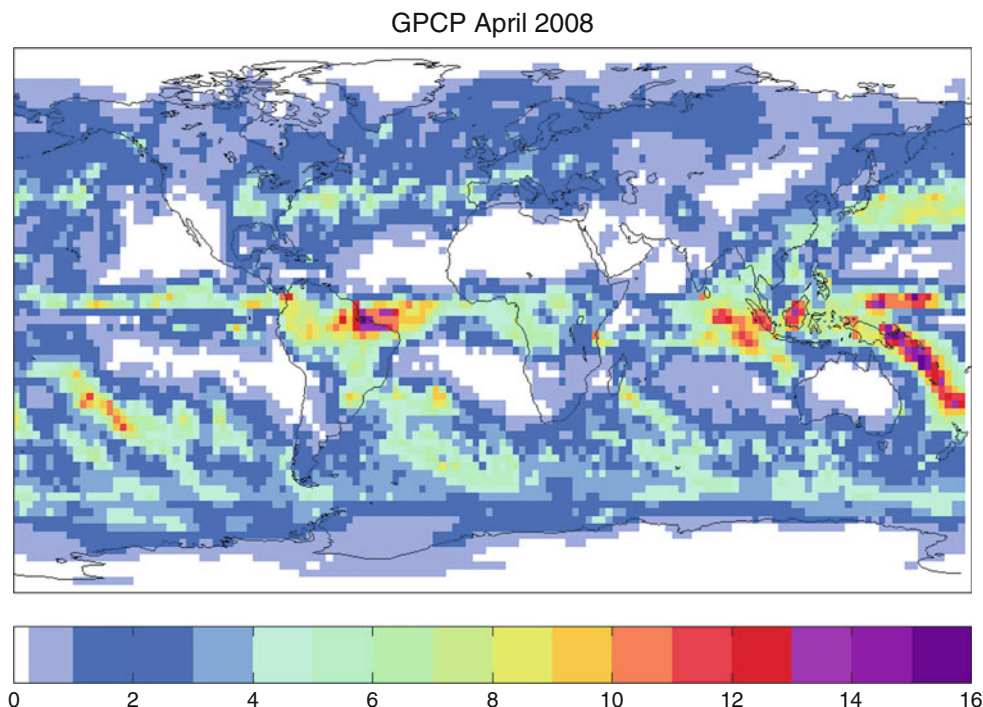
polar-orbiting passive microwave (PMW) or geosynchronous infrared (IR), where the more direct methods of the former generally give superior estimates. Additionally, the NASA Tropical Rainfall Monitoring Mission (TRMM; Kummerow et al., 1998) Precipitation Radar (Iguchi et al., 2000) also provides precipitation estimates, but these are limited to the tropics and the temporal sampling is insufficient for climate monitoring or prediction at the present time.

Despite the advantage of global coverage, satellite-only precipitation estimates are generally considered inferior to combinations of gauge data. Therefore, the most widely used precipitation datasets are combinations of multiple types of data. The two most commonly used precipitation datasets are the Global Precipitation Climatology Project (GPCP; Huffman et al., 1997; Adler et al., 2003) and the CPC Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1997). Both datasets are global, monthly, 2.5° combinations of IR, PMW, and gauge data and a single month of GPCP is shown in [Figure 1](#). A significant barrier to the use of these datasets for real-time monitoring is that the gauge components are not immediately available. The suite of GPCP outputs therefore includes a provisional product which contains most of the combined satellite data and a slightly different, less comprehensive gauge analysis which is available in near real time, but with slightly reduced accuracy which is suitable for climate monitoring. As well as use for the general monitoring of precipitation, these large-scale estimates can also be used in derived products as a key predictor. One particularly useful application is in public health where precipitation has been used for disease risk mapping through disease vector prediction (e.g., Lobitz et al., 2000).

Before the launch of the TRMM mission, there were insufficient polar-orbiting PMW satellites to provide better than daily temporal resolution, but this sampling has been greatly improved with the addition of TRMM (which includes the TRMM Microwave Imager) along with several other PMW instruments such as the Advanced Microwave Sounding Unit (AMSU; Weng et al., 2003) and the Advanced Microwave Scanning Radiometer (AMSR; Wilheit et al., 2003). Several datasets have merged the available PMW estimates with the lower quality but more frequently sampled IR data. These datasets have near-global coverage with resolutions as fine as 8 km every half hour (CMORPH; Joyce et al., 2004) and can include gauge data through a monthly bias correction (TRMM Multi-satellite Precipitation Analysis; TMPA; Huffman et al., 2007). These high-resolution precipitation products are used in a number of climate monitoring situations for near-term predictions such as flash flood prediction (Gupta et al., 2002; Harris et al., 2007) and landslide prediction (Hong et al., 2006, 2007).

Example 4: vegetation

The most commonly used remotely sensed vegetation product is the normalized difference vegetation index



Climate Monitoring and Prediction, Figure 1 GPCP V2 mean precipitation for April 2008, in mm day^{-1} .

(NDVI; Tucker, 1979) (see Vegetation). NDVI is based on near-infrared (NIR) and Red channels and is simply defined as $(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$ which is normalized to range from -1 to 1 . NDVI is well correlated with the health of vegetation and is based on how green the vegetation is within each observed pixel. Most surfaces equally absorb or reflect in the Vis and NIR channels and give a near zero value of NDVI. Chlorophyll, however, absorbs in the Vis channel while leaves reflect in the NIR channel so that large differences usually denote healthier (or more green) vegetation within a pixel. The most commonly used instrument for the estimation of NDVI has been the Advanced Very High resolution Radiometer (AVHRR) which has flown aboard several missions and continues to provide high-resolution data with near-global coverage. NDVI estimates are used in agricultural and drought monitoring where the health of the crops and vegetation is the primary monitoring target. An important attribute of satellite remote sensing is that observations transcend traditional international boundaries. This is an important characteristic for drought and crop monitoring, as NDVI products can be used to observe countries where ground observations are not available, but where food aid might be required. Derived products are also available, and systems have been developed which forecast disease outbreaks based on NDVI and other variables (Linthicum et al., 1999; Hendrickx et al., 2001; Anyamba et al., 2002).

In general, these studies are based on the use of NDVI and the health of plants as a proxy for the health of disease vectors.

Example 5: sea ice

Observations of sea ice extent and concentration are required for both climate change studies and seasonal monitoring (see [Sea Ice Concentration and Extent](#)). The cryosphere is known to be particularly sensitive to changes in global temperature, and the signs of climate change have been most visible in the diminishing Arctic sea ice and the alarming and spectacular breakup of Antarctic ice sheets. However, sea ice also plays an important role in several other climate monitoring datasets, and monitoring products are commonly used in a range of other remotely sensed products to screen out areas where retrievals are either not possible or require modification. Sea ice extent is estimated using a number of different techniques based on passive microwave and infrared observations as well as some limited in situ observations. The NOAA/NESDIS Interactive Multisensor Snow and Ice Mapping System (IMS; Helfrich et al., 2007) is a commonly used merged satellite sea-ice extent product. A single day of the IMS snow (white) and ice (light gray) product is shown in [Figure 2](#). Snow and ice cover are determined by a human forecaster based on remotely sensed data from all sources. Such products are useful

IMS 24km 15 April 2008



Climate Monitoring and Prediction, Figure 2 IMS Northern Hemisphere snow and ice cover (at 24 km resolution) for 15th April 2008.

for precipitation products where retrievals are often not possible in the presence of snow/ice and ice masks are sometimes used to remove erroneous values. Analyses of sea-surface temperature (SST), such as the Reynolds and Smith (1994) analysis discussed later on, use sea ice products to infer temperature over ice-covered surfaces.

As well as ice extent (or snow cover), the concentration or thickness of ice is of interest, as is the age of the ice (whether the ice is new or has existed for multiple years). Sea ice concentration can be inferred from the Quick Scatterometer (QuikSCAT), the Special Sensor Microwave/Imager (SSM/I), the Scanning Multichannel Microwave Radiometer (SMMR), and, more recently, the Advanced Microwave Scanning Radiometer (AMSR). Such measurements are critically important for climate change monitoring such as the estimation of the depletion of sea ice (e.g., Nghiem et al., 2006). More recently, the Gravity Recovery and Climate Experiment (GRACE) has been used to estimate ice sheet mass inferred from measurements of Earth's gravity.

Example 6: sea surface temperature

Estimates of [sea surface temperature](#) (SST) are used in both a climate change monitoring and a seasonal climate monitoring capacity (see [Sea Surface Temperature](#)). For climate change, estimates of SST are important as an indicator of observed changes but also for understanding the global response to carbon dioxide. In a monitoring sense, global SST estimates serve a wide range of functions.

One of the most important indicators of global climate is the El Niño/Southern Oscillation (ENSO) which is chiefly associated with changes in the location and extent of the Tropical-Pacific warm pool and is hence often defined as an average of sea surface temperatures. ENSO variability is associated with a broad range of global weather phenomena and is a dominant component of many climate predictions (further discussion appears later in this section). SST is also an important consideration in hurricane monitoring and prediction since warmer SSTs are required for the initiation of convection and are thus a key component for predicting areas where strengthening of tropical storms is more likely. One of the most widely used gridded SST analyses is the Reynolds and Smith (1994) optimal interpolation (OI; also known as Kriging) of available ship observations and satellite estimates. A single month of this SST estimate is shown in [Figure 3](#). The original dataset used only estimates from the five-channel Advanced Very High Resolution Radiometer (AVHRR) instrument which became operational aboard NOAA-7 in 1981, although it has been updated in successive years. As with many other parameters, merged IR/PMW estimates of SST are quickly becoming the standard and allow for superior estimates. One such dataset is the high-resolution version of the Reynolds and Smith (1994) SSTs which is available at 0.25° daily resolution and uses remotely sensed data from the pathfinder AVHRR from 1981 onward and a mix of AVHRR and Advanced Microwave Scanning Radiometer (AMSR) after its launch in 2002 (Reynolds et al., 2007). SST data have been directly or indirectly used for a variety of monitoring purposes including monitoring of the thermohaline circulation (Latif et al., 2004), as an input for malaria prediction models (Thomson et al., 2005) and as an input for models of coral bleaching (Maynard et al., 2008).

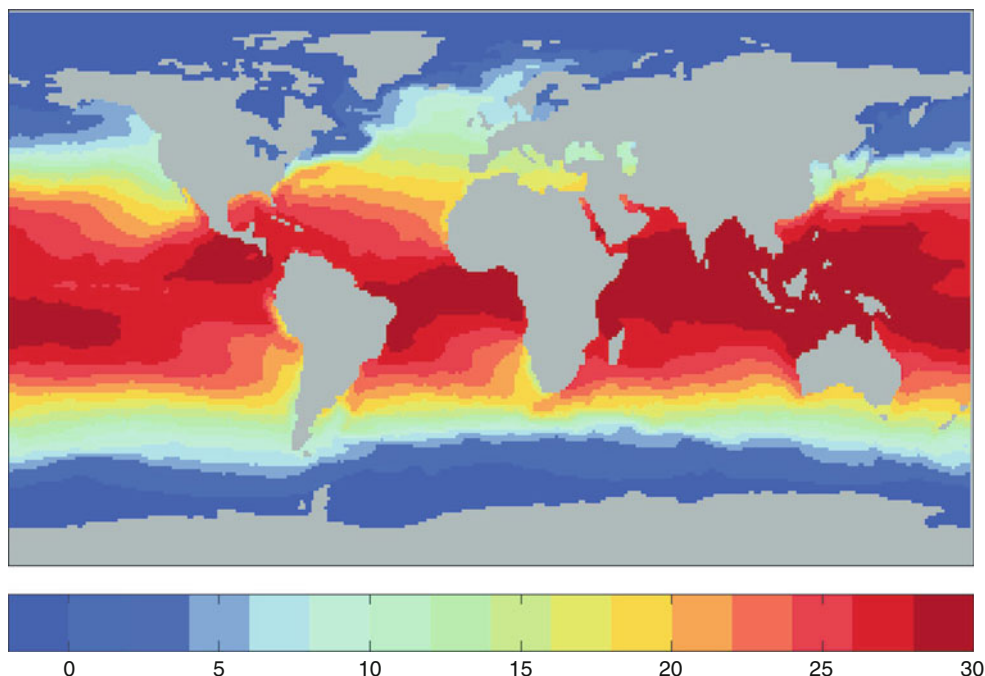
Climate prediction and reanalysis

Remotely sensed observations have become an important input to numerical weather and climate models as a main constituent of the analysis of the initial conditions. Numerical forecasts work by propagating a set of initial conditions according to the theoretical physics of the system. The process of estimating the initial condition is called *data assimilation* and involves collecting all of the inputs and merging them. A substantial part of the error in the model predictions is due to errors in the initial condition, and satellites have helped to reduce this error considerably through enhanced spatial and temporal sampling.

Reanalysis data

An issue common to all observations is that complete spatial and temporal coverage is not possible. In situ observations are frequently made at a point and are usually unevenly spaced in, at least, space. Remotely sensed observations tend to be more systematically sampled in space and time, but coverage is frequently not global and the sampling is often insufficient. Furthermore, satellite

OI V2 SST April 2008



Climate Monitoring and Prediction, Figure 3 NOAA OI SST V2 mean sea surface temperature for April 2008, in °C (Reynolds and Smith, 1994).

estimates are nearly always instantaneous whereas our interest for climate is in averages of the data. A polar-orbiting satellite might give a single measurement at a given location once each day, but climate studies require estimates for the whole day. In such cases, a model is often required to obtain estimates which are valid for the whole day.

In the last decade, model reanalysis data has become commonplace in climate science. Such reanalysis techniques are adapted from the data assimilation schemes of operational numerical [weather prediction models](#) and facilitate the combination of multiple observations of multiple parameters and the interpolation of these observations to data-sparse regions in four-dimensional space. Additionally, model forecasts (or nowcasts) can also be made to predict parameters not traditionally measured. Global reanalysis systems therefore provide rich new datasets which can provide data from anywhere in the four-dimensional system (within the model resolution constraints) based on all available in situ and remotely sensed observations. Current reanalysis is still somewhat limited in resolution due to constraints in processing capacity, but are ideal for a range of climate problems.

The first reanalysis to be widely used was the NCAR/NCEP reanalysis (Kalnay et al., 1996) which is still used for a range of purposes and is well suited for many climate monitoring purposes since it is available in near real time. Note that an updated version is now available, the NCEP-DOE AMIP-II Reanalysis (Kanamitsu et al., 2002), which

uses many of the same inputs but corrects a number of issues with the original dataset. The NCAR/NCEP reanalysis directly assimilated radiances from operational TOVS sounders once available (after 1979). Additional remotely sensed datasets were also indirectly used in the analysis as boundary fields. The Reynolds and Smith (1994) SST analysis was used after 1982 when AVHRR data became available and sea ice based on SMMR/SSM/I was used (as well as other non-remote sensing datasets). Another commonly used reanalysis is the European Centre for Medium Range [Weather Forecasting](#) (ECMWF) ERA-40 project which directly assimilated more satellite records than the NCAR/NCEP reanalysis. However, ERA-40 is not available in real time: Analyses from the ECMWF forecast assimilation scheme are available, but the ERA-40 product is not meant for climate monitoring.

The NCAR/NCEP reanalysis starts in 1946, and the ECMWF ERA-40 reanalysis starts in 1958, but much of the satellite data used in both assimilations starts in 1979. This discrepancy has led to discontinuities in the data (Bromwich and Fogt, 2004) which could lead to spurious trend estimates. More recent reanalysis products such as the Japanese JRA-25 (Onogi et al., 2007) and NASA's Modern Era Retrospective Reanalysis (MERRA; <http://gmao.gsfc.nasa.gov/research/merra/>) start in 1979 since remotely sensed observations of the atmosphere make this a far more data-rich period. At present, the ERA-40 reanalysis is not run operationally making it unsuitable for climate monitoring (although the ECMWF

does produce a similar assimilation for their operational model). Additionally, time required for data acquisition and processing make these data unsuitable for some monitoring activities.

Climate prediction

Whether physically or statistically based, predictions of the weather or climate are usually obtained by propagating a set of initial conditions forward to some given time point. In the case of weather forecasting (typically at scales less than 10 days), the focus is on the prediction of the most likely state (e.g., temperature) and individual weather events (e.g., precipitation) (see *Weather Prediction*). In contrast, climate prediction is focused on prediction of the mean state over the next few months, years, or decades. The bulk of all climate models is intended for prediction of climate change effect at the decadal or century scale (Randall et al., 2007). These models are not explored here, and we instead focus on models for seasonal predictions (between 10 days and several months).

Seasonal climate forecasts are made using one of two distinct approaches. The first is to use a dynamical, numerical model such as that used in operational models (see Goddard et al., 2001, for a review). Such models use a physical-based, dynamical representation of the atmosphere similar to those used for numerical weather forecasting. Forecasts are made by using simplified versions of the basic atmospheric and/or oceanic equations to propagate an initial field, usually based on observations. This initial field is largely based on remotely sensed observations, which are assimilated along with other, in situ observations. Seasonal climate predictions made with dynamical models are traditionally deterministic in nature, meaning that a single (hopefully optimal) answer is obtained. However, it is becoming increasingly common for ensemble approaches to be used whereby multiple realizations of the initial conditions are run through the model to assess the effect of measurement error in the initial conditions on the output. This is particularly useful for climate prediction since errors in the initial conditions are relatively small over a short period, but tend to be amplified for longer forecast lead times, an effect that can lead to forecasts being dominated by noise.

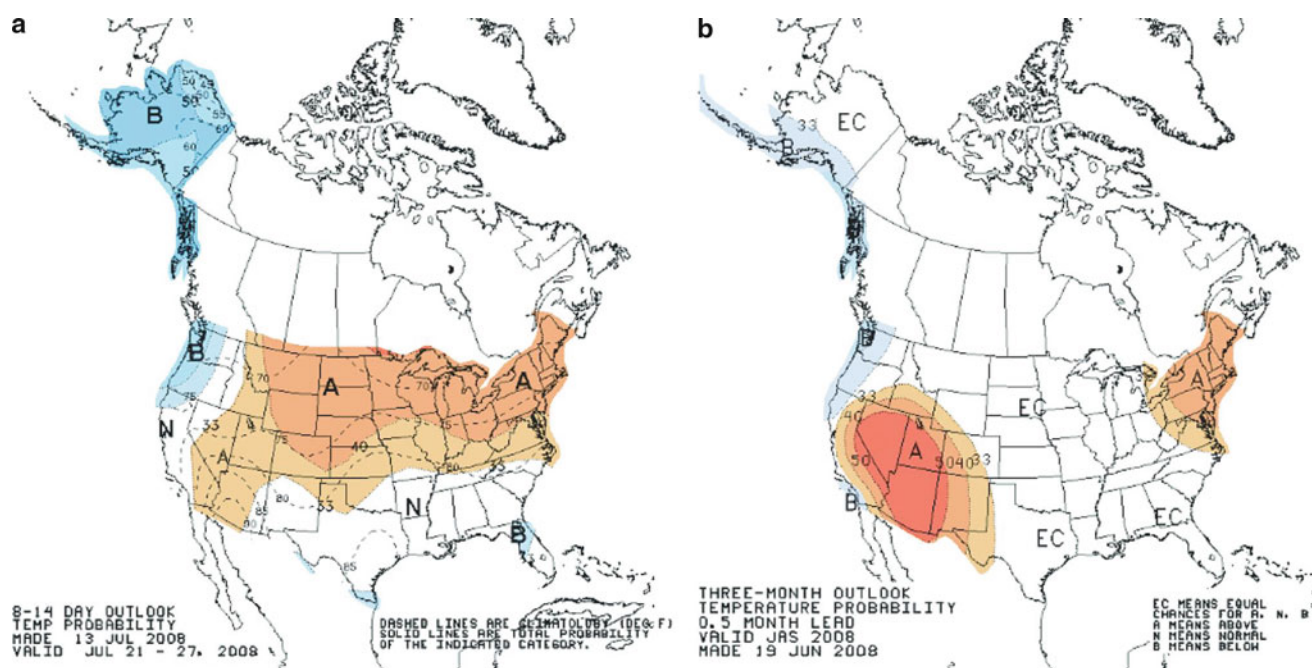
Another difference between weather and climate models is the treatment of the ocean and the inclusion of SSTs, which are intimately linked with atmospheric circulation and weather. SSTs change relatively slowly over time, and these changes are often too slow to make a large difference on weather scales. Weather models therefore tend to use only a dynamical atmosphere with prescribed ocean temperatures which are fixed in time. Within a season, changes in SST become important, and so many climate models (seasonal and decadal) use a dynamical ocean. Since the properties of the ocean and the atmosphere are different, it is common to use separate models and *couple* these models so that information passes between the two at set times. For instance, the ocean model

might be run for a day and then the new values used to update SSTs (or other parameters) for the atmospheric model, which in turn might provide wind estimates for the ocean model. Coupled models are an important tool for seasonal climate prediction that have become widely used at a variety of institutions (Stockdale et al., 1998; Graham et al., 2005; Saha et al., 2006; Luo et al., 2008).

The second commonly used approach for seasonal climate prediction is to use a statistical model to make forecasts based on relationships between some observed data and the quantity to be forecast. Statistical models partition the total variance in the data into that which can be explained by the predictor variables and a residual noise component which is modeled by some assumed error distribution. The advantage to this approach is that a full physical understanding of the system being modeled is not required since any unexplained noise is simply modeled in the residual. In practice, statistical prediction models are constructed by using some set of training parameters to develop a model and estimate the parameters of the model. Estimates of the response data are then generated from this model and standard errors (a measure of how incorrect the estimate is) can be generated based on the assumed distribution of the residual noise. Statistical approaches include a wide range of models from linear regression methods to neural networks and canonical correlation analysis. A common requirement of these models is that a lag relationship be known, whereby the predictors contain information about the future state of the response. Among other sources, climate indices such as ENSO and the North Atlantic Oscillation (NAO) have been found to give this lead information, and the ocean is another common input due to its thermal memory. This provides a logical contrast between statistical and dynamical models: Dynamical models rely on direct knowledge of the climate system to infer its future state, whereas statistical models require no knowledge of the system, instead relying on an indirect relationship between two or more variables.

Remotely sensed products play an important role in seasonal predictions as either a direct input to the models (dynamical or statistical) or in a monitoring capacity where they are used to infer the current mean state of the climate and its most likely tendency. Remotely sensed data is also used in a variety of statistical models. Some examples include models to predict rainfall (Nicholas and Battisti, 2008), SSTs (Landman and Mason, 2001), and prevalence of diseases in humans (Linthicum et al., 1999; Thomson et al., 2005) and plants (Boken et al., 2007). Forecasting of ENSO (Mason and Mimmack, 2002; Coelho et al., 2004) is also important since knowledge of its state helps forecasters infer many other weather characteristics and predictions can be used as the base for other forecasts. Additionally, some statistical models use forecast fields from climate models (which assimilate satellite data) to derive estimates (e.g., Hoshen and Morse, 2004; Thomson et al., 2006).

In practice, information from both statistical and dynamical prediction models is used along with the



Climate Monitoring and Prediction, Figure 4 CPC probability of above average, below average, or normal temperature for (a) 8–12 days and (b) 3 months for the United States. Downloaded from the NOAA/NCEP/CPC website: <http://www.cpc.ncep.noaa.gov/on> July 14th 2008.

assessment of the current state of the climate system, and seasonal forecast products are usually composites of multiple forecast and monitoring data. As with weather forecasts, a seasonal forecaster usually compiles the information into a single outlook based on the available products and some knowledge of the relative skill of each data source. This final product takes a variety of different forms, although it is common to have separate forecasts for the near and longer term (approximately 10–20 days and 3–6 months, respectively). Each product usually represents a time-integrated product over the period of interest rather than being an instantaneous value for several months time. One example of such a derived product is the CPC temperature probability product shown in Figure 4. Forecasts are shown for 8–12 days and 3 months as a probability of either above/below average or normal rather than an absolute value. Such metrics are common in seasonal climate prediction where the skill is insufficient to give much more than an indication of the likely state of the climate.

Conclusion

Remote sensing plays a large and varied role in climate monitoring and prediction, providing both the monitoring data and inputs for climate models (statistical or dynamical). At present, our ability to forecast climate is limited by our understanding of the global Earth system and the chaotic nature of the system. While the latter is unlikely to change, it should not be overlooked that current and future research satellite missions have the potential to

improve our understanding and enhance our ability to forecast climate further into the future and with greater accuracy.

Bibliography

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E., 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology*, 4(6), 1147–1167.
- Anyamba, A., Linthicum, K. J., Mahoney, R., Tucker, C. J., and Kelley, P. W., 2002. Mapping potential risk of rift valley fever outbreaks in African Savannas using vegetation index time series data. *Photogrammetric Engineering and Remote Sensing*, 68(2), 137–145.
- Bhartia, P., McPeters, R., Mateer, C., Flynn, L., and Wellemeyer, C., 1996. Algorithm for the estimation of vertical ozone profiles from the backscattered ultraviolet technique. *Journal of Geophysical Research*, 101(D13), 18793–18806.
- Boken, V. K., Hoogenboom, G., Williams, J. H., Diarra, B., Dione, S., and Easson, G. L., 2007. Monitoring peanut contamination in Mali (Africa) using AVHRR satellite data and a crop simulation model. *International Journal of Remote Sensing*, 29(1), 117–129.
- Bromwich, D. H., and Fogt, R. L., 2004. Strong trends in the skill of the ERA-40 and NCEP–NCAR reanalyses in the high and midlatitudes of the southern hemisphere, 1958–2001. *Journal of Climate*, 17, 4603–4619.
- Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weissenmayer, A., Richter, A., DeBeek, R., Hoogen, R., Bramstedt, K., Eichmann, K. U., Eisinger, M., and Perner, D., 1999. The global ozone monitoring experiment (GOME):

- mission concept and first scientific results. *Journal of the Atmospheric Sciences*, **56**, 151–175.
- Coelho, C. A. S., Pezzulli, S., Balmaseda, M., Doblas-Reyes, F. J., and Stephenson, D. B., 2004. Forecast calibration and combination: a simple Bayesian approach for ENSO. *Journal of Climate*, **17**, 1504–1516.
- de Toma, G., White, O. R., Chapman, G. A., and Walton, S. R., 2004. Solar irradiance variability: progress in measurement and empirical analysis. *Advances in Space Research*, **34**, 237–242.
- Forster, P., Ramaswamy, V., Artaxo, P., Bernsten, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., Van Dorland, R., Ramaswamy, V., Artaxo, P., Bernsten, T., Betts, R., Fahey, D. W., Haywood, J., Lean, J., Lowe, D. C., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Van Dorland, R., 2007. Changes in atmospheric constituents and in radiative forcing. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Fröhlich, C., 2007. Solar irradiance variability since 1978. *Space Science Reviews*, **125**(1–4), 53–65.
- Fröhlich, C., and Lean, J., 2004. Solar radiative output and its variability: evidence and mechanisms. *Astronomy and Astrophysics Review*, **12**, 273–320, doi:10.1007/s00159-004-0024-1.
- Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R., and Cane, M. A., 2001. Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology*, **21**, 1111–1152.
- Graham, R. J., Gordon, M., Mclean, P. J., Ineson, S., Huddleston, M. R., Davey, M. K., Brookshaw, A., and Barnes, R. T. H., 2005. A performance comparison of coupled and uncoupled versions of the Met Office seasonal prediction general circulation model. *Tellus A*, **57**(3), 320–339.
- Gupta, H., Sorooshian, S., Gao, X., Imam, B., Hsu, K., Bastidas, L., Li, J., and Mahani, S., 2002. The challenge of predicting flash floods from thunderstorm rainfall. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **360**(1796), 1363–1371.
- Harris, A., Rahman, S., Hossain, F., Yarborough, L., Bagtzoglou, A. C., and Easson, G., 2007. Satellite-based flood modeling using TRMM-based rainfall products. *Sensors*, **7**, 3416–3427.
- Helfrich, S. R., McNamara, D., Ramsay, B. H., Baldwin, T., and Kasheta, T., 2007. Enhancements to, and forthcoming developments in the interactive multisensor snow and ice mapping system (IMS). *Hydrological Processes*, **21**, 1576–1586.
- Hendrickx, G., Napala, A., Slingenberg, J. H. W., De Deken, R., and Rogers, D. J., 2001. A contribution towards simplifying area-wide Tsetse surveys using medium resolution meteorological satellite data. *Bulletin of Entomological Research*, **91**(5), 333–346.
- Hong, Y., Adler, R., and Huffman, G., 2006. Evaluation of the potential of NASA multi-satellite precipitation analysis in global landslide hazard assessment. *Geophysical Research Letters*, **33**, L22402, doi:10.1029/2006GL028010.
- Hong, Y., Adler, R., and Huffman, G., 2007. An experimental global prediction system for rainfall-triggered landslides using satellite remote sensing and geospatial datasets. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(6), 1671–1680.
- Hoshen, M. B., and Morse, A. P., 2004. A weather-driven model of malaria transmission. *Malaria Journal*, **3**, 32.
- Huffman, G. J., Adler, R. F., Arkin, P. A., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., and Schneider, U., 1997. The global precipitation climatology project (GPCP) combined precipitation dataset. *Bulletin of the American Meteorological Society*, **78**(1), 5–20.
- Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E. F., and Wolff, D. B., 2007. The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, **8**(1), 38–55.
- Iguchi, T., Kozu, T., Meneghini, R., Awaka, J., and Okamoto, K., 2000. Rain-profiling algorithm for the TRMM precipitation radar. *Journal of Applied Meteorology*, **39**, 2038–2052.
- Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, **5**(3), 487–503.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–471.
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M., and Potter, G. L., 2002. NCEP–DOE AMIP-II reanalysis (R-2). *Bulletin of the American Meteorological Society*, **83**, 1631–1643.
- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J., 1998. The tropical rainfall measuring mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, **15**, 809–817.
- Landman, W. A., and Mason, S. J., 2001. Forecasts of near-global Sea surface temperatures using canonical correlation analysis. *Journal of Climate*, **14**, 3819–3833.
- Latif, M., Roeckner, E., Botzet, M., Esch, M., Haak, H., Hagemann, S., Jungclaus, J., Legutke, S., Marsland, S., Mikolajewicz, U., and Mitchell, J., 2004. Reconstructing, monitoring, and predicting multidecadal-scale changes in the North Atlantic thermohaline circulation with Sea surface temperature. *Journal of Climate*, **17**, 1605–1614.
- Lin, Y., and Mitchell, K. E., 2005. The NCEP stage II/IV hourly precipitation analyses: development and applications. In *Preprints, 19th Conference on Hydrology*. San Diego, CA: American Meteorological Society, P1.2.
- Linthicum, K. J., Anyamba, A., Tucker, C. J., Kelley, P. W., Myers, M. F., and Peters, C. J., 1999. Climate and satellite indicators to forecast rift valley fever epidemics in Kenya. *Science*, **285**(5426), 397–400.
- Lobitz, B., Beck, L., Huq, A., Wood, B., Fuchs, G., Faruque, A. S. G., and Colwell, R., 2000. Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proceedings of the National Academy of Sciences of the United States of America*, **97**(4), 1438–1443.
- Luo, J. J., Masson, S., Behera, S. K., and Yamagata, T., 2008. Extended ENSO predictions using a fully coupled ocean–atmosphere model. *Journal of Climate*, **21**, 84–93.
- Mason, S. J., and Mimmack, G. M., 2002. Comparison of some statistical methods of probabilistic forecasting of ENSO. *Journal of Climate*, **15**, 8–29.
- Maynard, J. A., Turner, P. J., Anthony, K. R. N., Baird, A. H., Berkelmans, R., Eakin, C. M., Johnson, J., Marshall, P. A., Packer, G. R., Rea, A., and Willis, B. L., 2008. ReefTemp: an interactive monitoring system for coral bleaching using high-resolution SST and improved stress predictors. *Geophysical Research Letters*, **35**, L05603, doi:10.1029/2007GL032175.
- Mishchenko, M. I., Cairns, B., Kopp, G., Schueler, C. F., Fafaul, B. A., Hansen, J. E., Hooker, R. J., Itchkawich, T., Maring, H. B., and Travis, L. D., 2007. Accurate monitoring of terrestrial aerosols and total solar irradiance: introducing the glory

- mission. *Bulletin of the American Meteorological Society*, **88**, 677–691.
- Nghiem, S. V., Chao, Y., Neumann, G., Li, P., Perovich, D. K., Street, T., and Clemente-Colón, P., 2006. Depletion of perennial sea ice in the East Arctic Ocean. *Geophysical Research Letters*, **33**, L17501, doi:10.1029/2006GL027198.
- Nicholas, R. E., and Battisti, D. S., 2008. Drought recurrence and seasonal rainfall prediction in the Río Yaqui Basin. *Journal of Applied Meteorology and Climatology*, **47**, 991–1005.
- Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., and Taira, R., 2007. The JRA-25 reanalysis. *Journal of the Meteorological Society of Japan*, **85**, 369–432.
- Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R. J., Sumi, A., and Taylor, K. E., 2007. Climate models and their evaluation. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Reynolds, R. W., and Smith, T. M., 1994. Improved global sea surface temperature analyses using optimum interpolation. *Journal of Climate*, **7**, 929–948.
- Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G., 2007. Daily high-resolution blended analyses for sea surface temperature. *Journal of Climate*, **20**, 5473–5496.
- Rottman, G., 2006. Measurement of total and spectral solar irradiance. *Space Science Reviews*, **125**(1–4), 39–51.
- Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., Van den Dool, H. M., Pan, H. L., Moorthi, S., Behringer, D., Stokes, D., Peña, M., Lord, S., White, G., Ebisuzaki, W., Peng, P., and Xie, P., 2006. The NCEP climate forecast system. *Journal of Climate*, **19**, 3483–3517.
- Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S., and Balmaseda, M. A., 1998. Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392**, 370–373.
- Thomson, M. C., Mason, S. J., Phindela, T., and Connor, S. J., 2005. Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana. *The American Journal of Tropical Medicine and Hygiene*, **73**(1), 214–221.
- Thomson, M. C., Doblas-Reyes, F. J., Mason, S. J., Hagedorn, R., Connor, S. J., Phindela, T., Morse, A. P., and Palmer, T. N., 2006. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, **439**(7076), 576–579.
- Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**(2), 127–150.
- Velicogna, I., and Wahr, J., 2006. Acceleration of Greenland ice mass loss in spring 2004. *Nature*, **443**(7109), 329–331.
- Weng, F., Zhao, L., Poe, G., Ferraro, R., Li, X., and Grody, N., 2003. AMSU cloud and precipitation algorithms. *Radio Science*, **338**, 8068–8079.
- Wilheit, T. T., Kummerow, C., and Ferraro, R., 2003. Rainfall algorithms for AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 204–214.
- WMO (World Meteorological Organization), 2007. *Scientific assessment of ozone depletion: 2006*. Global Ozone Research and Monitoring Project – Report No. 50, Geneva: World Meteorological Organization, 572 pp.
- Woods, T. N., Eparvier, F. G., Fontenla, J., Harder, J., Kopp, G., McClintock, W. E., Rottman, G., Smiley, B., and Snow, M., 2004. Solar irradiance variability during the October 2003 solar storm period. *Geophysical Research Letters*, **31**, L10802, doi:10.1029/2004GL019571.
- Xie, P., and Arkin, P. A., 1997. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bulletin of the American Meteorological Society*, **78**(11), 2539–2558.

Cross-references

[Climate Data Records](#)
[Radiation, Solar and Lunar](#)
[Sea Surface Temperature](#)
[Stratospheric Ozone](#)
[Weather Prediction](#)

CLOUD LIQUID WATER

Fuzhong Weng

Center for Satellite Applications and Research (STAR),
 National Oceanic and Atmospheric Administration,
 College Park, MD, USA

Synonyms

CLW; LWP

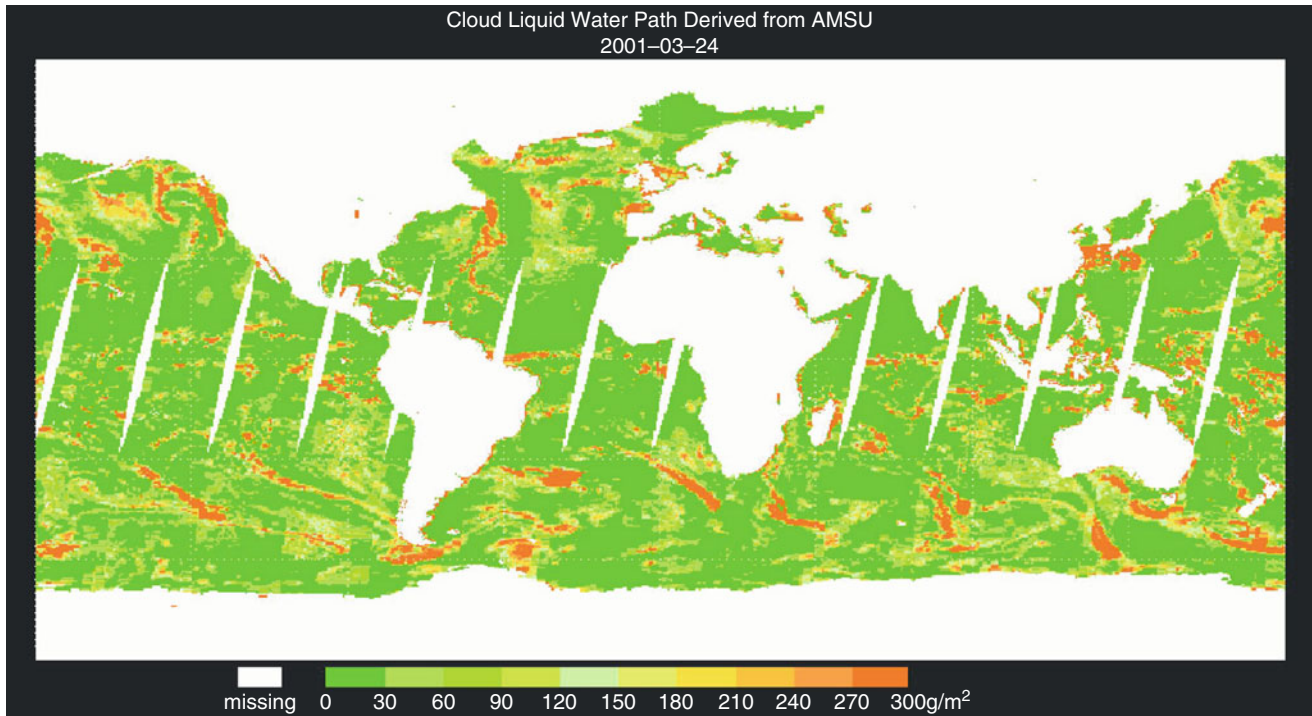
Definition

Total weight of cloud water in a vertical column of atmosphere for a unit of area.

Introduction

Clouds play a vital role in modulating climate and the earth's radiation budget. In the atmosphere, the latent heat release or consumption occurs either directly within the clouds or in the precipitation produced from inside the clouds. Clouds strongly affect the radiative fluxes through the atmosphere. Thus, the measurements of hydrometeor variables on clouds in various water phases critically affect the numerical weather prediction (NWP) simulation and climate modeling.

Global measurements of the cloud liquid water path can be determined by satellite-measured microwave brightness temperatures, which quantify the thermal emission of cloud particles. In the early 1970s, the feasibility of the microwave measurement of cloud liquid water was demonstrated by a Nimbus-6 scanning microwave spectrometer. A statistical relationship was first derived between the brightness temperatures at 23 and 31 GHz and cloud liquid water using Nimbus-6 scanning microwave spectrometer data (Grody, 1976). The large-scale distribution of cloud liquid water was obtained over the Pacific Ocean (Grody et al., 1980). This capability was further displayed by Nimbus-7 scanning multichannel microwave radiometer (SMMR) data (Takeda and Liu, 1987). However, more algorithms for cloud liquid water were developed for the Special Sensor Microwave Imager (SSM/I) flown on the defense meteorological



Cloud Liquid Water, Figure 1 Global cloud liquid water retrieved from NOAA-15 Advanced Microwave Sounding Unit (AMSU). Only satellite descending node data are used in this retrieval. Land and sea ice areas are shown as missing because of the incapability of calculating surface emissivity.

satellite program (e.g., Alishouse et al., 1990; Greenwald et al., 1993; Liu and Curry, 1993; Weng and Grody, 1994; Wentz, 1997). The algorithm was further refined for Advanced Microwave Sounding Unit (AMSU) measurements at 23.8, 31.4 GHz (Weng et al., 2003).

Algorithm theoretical database

Cloud liquid water is also referred as the liquid water path. It can be measured from both passive and active remote sensing technology deployed in space and on the ground. From satellites, brightness temperatures at lower microwave frequencies directly respond to the emission signals from clouds and raindrops. Over oceans, the brightness temperatures first increase and then become saturated and decrease as cloud liquid water increases. The nonlinear response is a result of emission and scattering from both small cloud droplets and large raindrops. The saturation point normally varies with frequency and is, for example, 8, 3, and 1 kg/m² (mm) at 10.65, 18.7, and 36.5 GHz, respectively, for a typical warm rain situation. Since the actual liquid water path ranges within several millimeters, it is necessary to use a composite algorithm (Weng and Grody, 1994) to retrieve cloud liquid water to cover the range from non-raining to raining clouds.

In the absence of scattering from precipitation, brightness temperature at a microwave frequency can be derived as the function (Weng et al., 2003):

$$TB = T_s [1 - (1 - \varepsilon)T^2], \quad (1)$$

where ε and T_s are the surface emissivity and surface temperature, respectively, and the atmospheric transmittance is:

$$T = \exp[-(\tau_o + \tau_v + \tau_l)/\mu], \quad (2)$$

where τ_o is the optical thicknesses of oxygen. The optical thicknesses of cloud and water vapor are expressed as $\tau_L = \kappa_L L$ and $\tau_v = \kappa_v V$, respectively, where the liquid water path is $L = \int_0^{\infty} w(z) dz$ and the water vapor path is $V = \int_0^{\infty} \rho_v dz$. The cloud mass absorption

coefficient κ_l is approximated as $\kappa_l = \frac{6\pi}{\lambda \rho_w} \text{Im} \left\{ \frac{m^2 - 1}{m^2 + 2} \right\}$ through Rayleigh's approximation.

Equation 1 provides a fundamental theory for microwave remote sensing of both atmospheric liquid water and water vapor from space over oceanic conditions. In general, at least two frequencies are required, with one being more sensitive to liquid and the other to water vapor. Note that for land conditions

where the emissivity is normally high (typically greater than 0.9), brightness temperature decreases as cloud liquid water increases. The depression from non-raining clouds is also typically very small, less than several degrees Kelvin. Thus, it is difficult to detect the liquid-phase clouds over land where its emissivity is high and variable.

Using two channel measurements at 23.8 and 31.4 GHz from the Advanced Microwave Sounding Unit (AMSU), TB_{23} and TB_{31} , Weng et al. (2003) derived

$$V = a_0 \mu [\ln(T_s - TB_{31}) - a_1 \ln(T_s - TB_{23}) - a_2], \quad (3)$$

and

$$V = b_0 \mu [\ln(T_s - TB_{31}) - b_1 \ln(T_s - TB_{23}) - b_2], \quad (4)$$

respectively, where the coefficients, a_0 and b_0 , are functions of cloud and water vapor mass absorption coefficients, and $a_{1,2}$ and $b_{1,2}$ are functions of surface emissivity and surface temperature, respectively.

The figure below displays a global distribution of cloud liquid water over oceans derived from the use of the AMSU 23.8 and 31.4 GHz measurements. Note that the AMSU measurements during a 24 h period from its descending node do not completely cover the globe because of the presence of orbital gaps. Also, the retrievals are not performed over land, snow, and sea ice conditions due to large emissivity variations. The unit of cloud liquid water is g/m^2 . Note that low clouds over oceans to the west coast of South America are detected very well and their liquid water path is on the order of 100 g/m^2 . The weather systems having relatively high amount of cloud liquid water are associated with those frontal systems and the clouds within the inter-tropical convergence zone (ITCZ) (Fig. 1).

Summary

Satellite passive microwave measurements provide accurate retrievals of cloud liquid water over oceans. It still remains difficult to retrieve cloud liquid water over land from the emission-based algorithms. Further studies will focus on extending the retrievals with more advanced radiative transfer schemes that include scattering from clouds and precipitation. Also, different algorithms should be developed for high and variable emissivity conditions which are typical of land, snow, and sea ice conditions.

Bibliography

- Alishouse, J. C., Alishouse, J. C., Snider, J. B., Westwater, E. R., Swift, C. T., Ruf, C. S., Snyder, S. A., Vongsathorn, J., and Ferraro, R. R., 1990. Determination of cloud liquid water content using the SSM/I. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 817–822.
- Greenwald, T. J., Stephens, G. L., and Vonder Haar, T. H., 1993. A physical retrieval of cloud liquid water over the global oceans using special sensor microwave/imager (SSM/I) observations. *Journal of Geophysical Research*, **98**, 18471–18488.
- Grody, N. C., 1976. Remote sensing of atmospheric water content from satellite using microwave radiometry. *IEEE Transactions on Antennas and Propagation*, **24**, 155–162.

- Grody, N. C., Gruber, A., and Shen, W. C., 1980. Atmospheric water content over tropical Pacific derived from the Nimbus-5 Scanner Radiometry. *Journal of Applied Meteorology*, **8**, 986–996.
- Liljegren, J. C., Clothiaux, E. E., Mace, G. G., Mace, G. G., Kato, S., and Dong, X., 2001. A new retrieval of cloud liquid water path using a ground-based microwave radiometer and measurements of cloud temperature. *Journal of Geophysical Research*, **106**, 14485–14500.
- Liu, G., and Curry, J. A., 1993. Determination of characteristic features of cloud liquid water from satellite microwave measurements. *Journal of Geophysical Research*, **98**, 5069–5092.
- Takeda, T., and Liu, G., 1987. Estimation of atmospheric liquid water amount by Nimbus 7 SMMR data: a new method and its application to the western north Pacific region. *Journal of the Meteorological Society of Japan*, **65**, 931–946.
- Weng, F., and Grody, N. C., 1994. Retrieval of cloud liquid water over oceans using special sensor microwave imager (SSM/I). *Journal of Geophysical Research*, **99**, 25535–25551.
- Weng, F., Zhao, L., Ferraro, R., Poe, G., Li, X., and Grody, N., 2003. Advanced microwave sounding unit cloud and precipitation algorithms. *Radio Science*, **38**, 8086–8096.
- Wentz, F. J., 1997. A well-calibrated ocean algorithm for special sensor microwave/imager. *Journal of Geophysical Research*, **102**, 8703–8718.

CLOUD PROPERTIES

Matthew Lebsock and Steve Cooper
Jet Propulsion Laboratory, California Institute of
Technology, Pasadena, CA, USA

Definitions

Droplet size distribution. The number distribution of water droplets as a function of droplet size.

Cloud optical depth. A unitless measure of the column integrated radiative extinction of a cloud.

Water content. The liquid or ice water content of an atmospheric volume.

Water path. The vertical integral of the cloud water content.

Droplet number concentration. The number of droplets within an atmospheric volume.

Droplet effective radius. A characteristic droplet size defined as the third moment of the droplet size distribution divided by the second moment.

Radar. Radio detection and ranging.

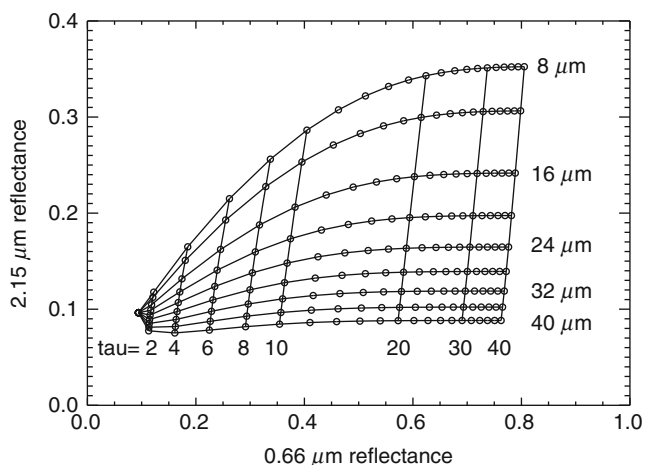
Lidar. Light detection and ranging.

Polarization. A property of electromagnetic radiation that describes the orientation and phase of its oscillations.

Brightness temperature. The blackbody temperature that is inferred from a measured spectral intensity.

Introduction

This entry reviews common remote sensing methods to infer cloud properties. Knowledge of cloud properties is of primary interest because they (1) have a large influence on the reflection and absorption properties of the



Cloud Properties, Figure 1 Relationship between visible $0.66\ \mu\text{m}$ and near-infrared $2.15\ \mu\text{m}$ reflectances for water clouds with optical depths ranging from 0 to 40 and effective radii ranging from 8 to $40\ \mu\text{m}$ for a case with overhead sun and nadir viewing angle.

atmosphere, thus having a profound effect on the Earth's energy balance, and (2) are related to the turbulent and microphysical processes that govern the lifecycle of clouds and the formation of rainfall, thus influencing the cycling of water within the Earth system.

The overarching goal of cloud property remote sensing is to characterize the droplet size distribution of a cloud. The droplet size distribution is highly variable in space and time, and it is therefore common to approximate its character by the integrated water content and the droplet effective radius, which can then be related to analytic functions such as the log-normal distribution or the gamma distribution. The focus here is on satellite remote sensing of cloud microphysical properties; however, it is noted that similar techniques are used to infer cloud properties from ground-based and airborne platforms as well.

Visible and near-infrared techniques

Visible and near-infrared techniques have been developed to simultaneously estimate the cloud optical depth and droplet effective radius from measurements of reflected sunlight in a visible and near-infrared channel (Nakajima and King, 1990). The physical basis for these methods lies in the fact that reflected visible radiation is primarily sensitive to the cloud optical depth, whereas the reflection of near-infrared radiation is largely sensitive to the droplet effective radius. Figure 1 shows an example of these dependencies for liquid clouds. Increasing reflectance in the visible channel corresponds to increasing cloud optical depth, whereas increasing reflectance in the near-infrared channel corresponds to decreasing cloud effective radius. The cloud liquid water path may then be derived from the cloud optical depth and effective radius (Stephens, 1978). Analogous methods are used for the

remote sensing of ice clouds. The visible and near-infrared methods are typified by the MODerate resolution Imaging Spectroradiometer (MODIS) cloud retrievals (Platnick et al., 2003). A notable limitation of visible and near-infrared methods is that they rely on reflected sunlight and therefore may only be performed during sunlit hours.

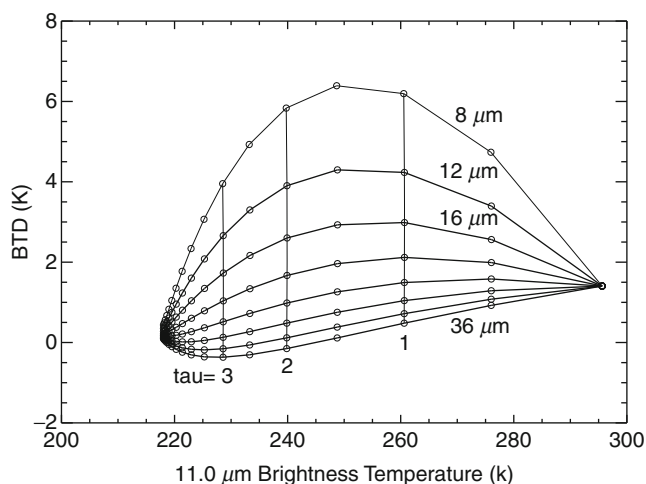
Microwave techniques

Passive measurements of microwave brightness temperatures have been used to estimate the cloud water path, thus providing a useful integral constraint on the vertical distribution of liquid water content. The physics governing these measurements is that at microwave frequencies, the emission by cloud water may be approximated by the Rayleigh limit (Greenwald, 2009), under which emission is linearly related to cloud water mass. Channels centered near 37 GHz demonstrate the largest signal-to-noise characteristics and dynamic range for the remote sensing of cloud liquid water.

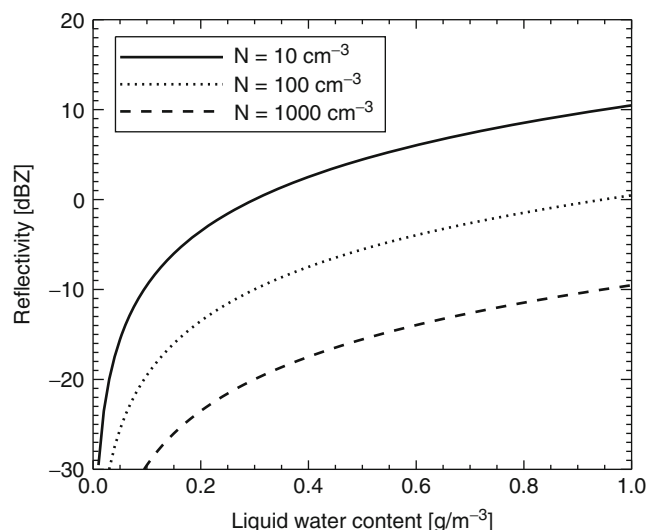
A complication of passive microwave techniques is that the brightness temperatures are also sensitive to the atmospheric temperature and water vapor content as well as the surface temperature and emissivity. Sensitivity to these parameters varies with frequency and measurement polarization. Therefore, passive microwave satellite instruments generally make observations at several frequencies and polarization states, allowing for the simultaneous determination of these properties. The algorithm of Hilburn and Wentz (2008) typifies this approach. A further complication of passive microwave estimates of cloud water path is that they are only possible over oceanic backgrounds where the surface emission is small. The emission from land surfaces is generally large and dominates the much smaller atmospheric emission signal.

Infrared techniques

An infrared method known as the split-window technique exploits differences in radiative properties of cloud particles at two wavelengths in the atmospheric window to estimate ice cloud properties from satellite-observed brightness temperatures. Small particles less than about $30\ \mu\text{m}$, for example, absorb radiation more efficiently at a wavelength of $12\ \mu\text{m}$ than at $11\ \mu\text{m}$. Clouds composed of small particles therefore appear colder at $12\ \mu\text{m}$ than at $11\ \mu\text{m}$ from a satellite perspective. Such radiometric signatures were first exploited by Inoue (1985) to detect thin cirrus. Since the ratio of absorption at the two wavelengths is a function of particle size, the technique can also be used to infer cloud effective radius from the brightness temperature difference (ΔT_b) between the two wavelengths (Prabhakara et al., 1988). Infrared radiative transfer (see entry *Radiative Transfer, Theory*) calculations for a set of cloud properties resemble an “arch,” as seen in Figure 2. The right foot of the arch corresponds to the clear-sky emitting temperature of the atmosphere, while the left foot represents an optically thick cloud where observed brightness temperatures approach the



Cloud Properties, Figure 2 Relationship between TB (11 μm) and BTD (11–12 μm) for cirrus clouds with optical depths ranging from 0 to 10 and effective radii ranging from 8 to 32 μm . The clouds are composed of randomly oriented randomized hexagonal ice aggregates (Baran et al., 2003) and have a temperature of 217 K.



Cloud Properties, Figure 3 Theoretical relationship between liquid water content and reflectivity for three different values of number concentration (N) using the equations of Matrosov et al. (2004).

cloud thermodynamic temperature. Intermediate values provide information on cloud properties. Cloud optical depth is found from T_{b11} along the x-axis and cloud effective radius from ΔT_b along the y-axis. Since the technique has sensitivities limited to thin clouds and small particles, it has been applied infrequently at the operational scale. Heidinger and Pavolonis (2009), however, produced a global, multi-decadal set of cloud properties from the split-window technique based upon the AVHRR Pathfinder Atmospheres Extended dataset.

Visible polarization techniques

The polarization state of reflected visible radiation can be used to determine the effective radius of liquid clouds (Bréon and Goloub, 1998). The physical basis for this technique lies in the observation that at scattering angles between 150° and 170° liquid clouds produce cloud-bow features in the polarized reflection. The angular and spectral characteristics of these cloud-bows may be used to determine the liquid cloud effective radius quite accurately. Use of this technique is limited to sunlit hours over large homogenous cloud layers and is therefore not globally applicable. Despite its limited applicability, it has been used to identify high biases of approximately 2μ in effective radius estimates from the less accurate visible and near-infrared techniques (Bréon and Doutriaux-Boucher, 2005).

Active techniques

The successful flight of the CloudSat cloud profiling radar and the CALIOP lidar has permitted active techniques for the remote sensing of cloud properties from space. These instruments transmit discrete pulses of radiation and then

measure the backscattered signal as a function of time from transmission resulting in a vertical profile of backscattered reflectivity. A major advantage of active methods over passive methods is that they provide precise cloud boundaries and observed vertical profiles of the cloud microphysical quantities.

Radar profiling of cloud water content relies on either physical or empirical relationships between the cloud water content or cloud effective radius and the backscattered reflectivity that take the form of a power law (Matrosov et al., 2004). Figure 3 shows some relationships between reflectivity and cloud water content. Note the wide distribution of relationships between the cloud water content and the reflectivity, which results in a large uncertainty in the cloud property retrievals without the addition of an integral constraint. Integral constraints that have been proposed include the cloud optical depth (Austin and Stephens, 2001) or an ancillary passive microwave observation (Dong and Mace, 2003).

Detection of clouds with cloud radar is limited by the radar minimum detectable signal. Clouds that are too thin or composed of small droplets can create reflectivities that lie below the minimum detectable signal. However, the strong lidar backscattering signal permits detection of these clouds. This situation is particularly common with thin cirrus and stratus clouds. Methods analogous to those using radar have been developed to estimate ice cloud microphysical properties from the lidar backscatter when the radar signal falls below the minimum detectable signal (Delanoe and Hogan, 2008). These lidar methods are not feasible for water clouds because of the strong attenuation of the lidar beam by liquid water.

Summary

This entry reviews the primary methods by which cloud microphysical properties have been estimated from satellite remote sensing. The common methods covered include visible and near-infrared, passive microwave, infrared, visible polarization, and active techniques. Each of these techniques has their own strengths and weaknesses and offers a unique piece of information regarding cloud properties. A topic of current research in cloud remote sensing is understanding the synergies between these various methods and determining how they might best be combined to more accurately estimate cloud properties.

Acknowledgments

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Austin, R. T., and Stephens, G. L., 2001. Retrieval of stratus cloud microphysical parameters using millimeter-wave radar and visible optical depth in preparation for CloudSat 1. Algorithm formulation. *Journal of Geophysical Research*, **106** (D22), 28233–28242.
- Baran, A., Havemann, S., Francis, P., and Watts, P., 2003. A consistent set of single-scattering properties for cirrus cloud: tests using radiance measurements from a dual-viewing multi-wavelength satellite-based instrument. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **79**, 549–567.
- Br  on, F.-M., and Doutriaux-Boucher, M., 2005. A comparison of cloud droplet radii measured from space. *IEEE Transactions on Geoscience and Remote Sensing*, **43**(8), 1796–1805.
- Br  on, F., and Goloub, P., 1998. Cloud droplet effective radius from spaceborne polarization measurements. *Geophysical Research Letters*, **25**(11), 1879–1882.
- Delanoe, J., and Hogan, R. J., 2008. A variational scheme for retrieving ice cloud properties from combined radar, lidar, and infrared radiometer. *Journal of Geophysical Research*, **113**, D07204, doi:10.1029/2007JD009000.
- Dong, X., and Mace, G. G., 2003. Profiles of low-level stratus cloud microphysics deduced from ground-based measurements. *Journal of Atmospheric and Oceanic Technology*, **20**, 42–53.
- Greenwald, T. J., 2009. A 2 year comparison of AMSR-E and MODIS cloud liquid water path observations. *Geophysical Research Letters*, **36**, L20805.
- Heidinger, A. K., and Pavolonis, M. J., 2009. Gazing at cirrus clouds for 25 years through a split window. Part I: methodology. *Journal of Applied Meteorology and Climatology*, **48**, 1100–1116.
- Hilburn, K. A., and Wentz, F. J., 2008. Intercalibrated passive microwave rain products from the unified microwave ocean retrieval algorithm (UMORA). *Journal of Applied Meteorology and Climatology*, **47**, 778–794.
- Inoue, T., 1985. On the temperature and effective emissivity determination of semi-transparent cirrus clouds by bispectral measurements in the 10 micron window region. *Journal of the Meteorological Society of Japan*, **63**, 88–99.
- Matrosov, S. Y., Uttal, T., and Hazen, D. A., 2004. Evaluation of radar reflectivity–based estimates of water content in stratiform marine clouds. *Journal of Applied Meteorology*, **43**, 405–419.
- Nakajima, T., and King, M. D., 1990. Determination of the optical thickness and effective particle radius of clouds from reflected

solar radiation measurements. Part I: theory. *Journal of the Atmospheric Sciences*, **47**, 1878–1893.

- Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riedi, J. C., and Frey, R. A., 2003. The MODIS cloud products: algorithms and examples from Terra. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 459–473.
- Prabhakara, C., Fraser, R. S., Dalu, G., Wu, M. C., Curran, R. J., and Styles, T., 1988. Thin cirrus clouds: seasonal distribution over oceans deduced from nimbus-4 IRIS. *Journal of Applied Meteorology*, **27**, 379–399.
- Stephens, G., 1978. Radiation profiles in extended water clouds. II: parameterization schemes. *Journal of Atmospheric Sciences*, **35**, 2123–2132.

Cross-references

[Cloud Liquid Water](#)
[Lidar Systems](#)
[Microwave Radiometers](#)
[Radars](#)
[Rainfall](#)
[Reflected Solar Radiation Sensors, Polarimetric](#)
[Water Vapor](#)

COASTAL ECOSYSTEMS

Xiaojun Yang
 Department of Geography, Florida State University,
 Tallahassee, FL, USA

Synonyms

Coastal zone; Littoral ecosystems

Definition

Coastal ecosystems are typically found at a physical region extending from the edge of the continental shelf to the intertidal and nearshore terrestrial area. They include saline, brackish, and freshwaters, as well as coastlines and the adjacent lands that can extend to the entire coastal watershed. Globally, coastal ecosystems comprise a wide array of nearshore terrestrial, intertidal, benthic, and pelagic marine ecosystems.

Introduction

Coastal ecosystems, by virtue of their position at the interface between truly terrestrial and oceanic ecosystems, belong to the most dynamic and productive ecosystems on Earth. They are among the most important ecosystems, providing numerous ecological, economic, cultural, and aesthetic benefits and services. Although they comprise only 20 % of all land area, coastal areas are now the home of nearly half of the global population (Burke et al., 2001). Increased coastal population and intense development threaten and degrade global coastal ecosystems, placing an elevated burden on organizations responsible for the planning and management of these sensitive areas (Hinrichsen, 1998; Hobbie, 2000; National Research Council, 2000; Selman et al., 2008).

Coastal ecosystem management involves the procedure of monitoring which is based on a reliable information base. Conventional field-based mapping methods are still vital but often logistically constrained. Because of cost-effectiveness and technological soundness, remote sensing has increasingly been used to develop useful sources of information supporting decision making for various coastal applications (e.g., Yang et al., 1999; Yang, 2005a, b, 2008, 2009a, b). But coastal environments challenge the applicability and robustness of remote sensing because they exhibit extreme variations in spatial complexity and temporal variability (Klemaš, 2009). Encouragingly, recent innovations in data, technologies, and theories in the wider arena of remote sensing have provided scientists with invaluable opportunities to advance the studies on coastal environments.

Given the above context, this entry will review several major types of coastal ecosystems, followed by a discussion on how remote sensing can be used to characterize coastal waters, submerged aquatic vegetation, benthic habitats, coastal wetlands, and watersheds. It will finally highlight several areas that need further research and development.

Major coastal ecosystems

Comprehensive reviews on various coastal ecosystem types are given elsewhere (e.g., Mann, 2000; Beatley et al., 2002). Our current discussion targets several types of coastal ecosystems including estuaries, benthic systems, sea grass systems, coastal marshes, and mangroves, because they have been extensively studied by remote sensing. Note that these ecosystems are largely recognized according to the main primary producer with the exception of estuaries. As a rather large ecosystem, estuaries can include the other four as subsystems, but the latter may also occur outside an estuary.

Estuaries are partially enclosed bodies of coastal waters typically found where freshwater from rivers meet with saltwater from the ocean. They are often known as bays, lagoons, harbors, inlets, sounds, or fjords. Estuaries are important ecosystems, providing goods and services that are ecologically, economically, and culturally indispensable. Ecologically, estuaries are not only the “nurseries of the sea,” providing habitats for many marine organisms, but also serve as a natural buffer that filters out much of the sediments and pollutants carried in by terrestrial runoff, creating cleaner water that eventually benefits both human and marine life. Economically, estuaries support significant fisheries, tourism, and other commercial activities and the development of important public infrastructure, such as harbors and ports. Culturally, estuaries are often the focal points for recreation, commerce, scientific research and education, and aesthetic enjoyment.

The benthic system is the community of organisms living on the bottom of oceans in areas not colonized by macrophytes (Mann, 2000). Benthic habitats are virtually bottom environments with distinct physical, chemical, and

biological characteristics. They vary widely depending upon their location, depth, salinity, and sediment. Benthic habitats in areas with depth greater than 200 m have been much less commonly observed and mapped. Estuarine and nearshore benthic habitats can be highly diverse, including submerged mudflats, rippled sand flats, rocky hard-bottom habitats, shellfish beds, and coral reefs. Note that sea grass beds can be described as a benthic system, but they are treated separately (see above). In terms of ecological functioning, the benthic system serves as the site of nutrient regeneration and the site of considerable secondary production that is utilized by important predators, such as bottom-feeding fish and crustaceans (Mann, 2000).

Sea grass are aquatic flowering plants that live fully submerged in the saline coastal environment and are also called “submerged aquatic vegetation.” Sea grass can form extensive beds or meadows, dominated by one or more species. They are distributed worldwide in soft sediments from mean low tide level to the depth limit determined by the penetration of light that permits sea grass plants to photosynthesize. Sea grass beds are a highly diverse and productive ecosystem, and they almost always support more invertebrates and fish than the adjacent areas lacking sea grass (Mann, 2000). Sea grass beds provide coastal zones with a number of ecosystem goods and services, such as fishing grounds, wave protection, oxygen production, and protection against coastal erosion.

Coastal marshes are predominately grasslands that are periodically flooded by tides in the intertidal regions. The salinity from salt or brackish tidal waters creates a salt-stressed aquatic environment where halophytic plants thrive. Coastal marshes may be classified as salt marshes, brackish marshes, and freshwater tidal marshes. They may be associated with estuaries and are also along waterways between coastal barrier islands and the inner coast. Coastal marshes are of great ecological values because they serve as the nursery grounds for fish, habitats for a wide variety of wildlife, and the buffer zones to protect water quality.

The mangrove ecosystem is commonly found in tropical and subtropical tidelands throughout the world. The mangrove family of plants dominates this coastal wetland ecosystem due to their ability to thrive in the saline coastal environment. The three mangrove species commonly grown in the USA are red, black, and white mangroves. The term “mangroves” can narrowly refer to these species but most commonly refers to the habitat and entire plant assemblages. Because mangroves are constantly replenished with nutrients, they sustain a huge population of organisms that in turn feed fish and shrimp, which further support a variety of wildlife; their physical stability helps prevent shoreline erosion, shielding inland areas from damage during severe storms and waves.

Remote sensing of coastal ecosystems

Modern remote sensing technology began with the invention of the camera more than one century ago, and by now, a large number of remote sensing systems have

Coastal Ecosystems, Table 1 Characteristics of selected current and future satellite remote sensing systems relevant to coastal environments

Satellite/sensor	Date launched or planned	Spectral range and bands	Spatial resolution	Radiometric resolution (bit)	Revisit (day)	Swath width (km)	Specific coastal applications
OrbView-2/SeaWiFS	1 August 1997	0.402–0.885 μm 8 bands	1.1 km	10	1	2,800	Ocean color, phytoplankton, ocean carbon cycle
ENVISAT/MERIS	1 March 2002	0.390–1.040 μm 15 bands	300–1,200 m	12	3	1,150	Ocean color, phytoplankton, turbidity, sediment, red tides, vegetation
KOMPSAT-1/OSMI	20 December 1999	0.400–0.900 μm 6 bands	850 m	10	2	800	Ocean color, phytoplankton, turbidity
Terra–Aqua/MODIS (Terra)	18 December 1999 4 May 2002 (Aqua)	0.405–14.385 μm 16 bands	250–1,000 m	12	1	2,330	Sea surface temperature, ocean color, phytoplankton, turbidity, circulation, land cover and vegetation
NOAA POES (NOAA-14 to 18)/AVHRR	30 December 1994 (14) 13 May 1998 (15) 21 September 2000 (16) 24 June 2002 (17) 20 May 2005 (18) 21 November 2000	0.580–12.500 μm 6 bands	1.1 km	10	0.5–1	2,400	Sea surface temperature, turbidity, circulation, land cover and vegetation
EO-1/Hyperion	October 28, 2011 (S-NPP)	0.400–2.40 μm 220 bands	30 m	16	16	7.5 $\times$ 180	Bathymetry, vegetation, littoral processes
Suomi-NPP/VIIRS	October 28, 2011 (S-NPP)	0.412–12.013 μm 22 bands	370–740 m	12–14	0.5–1	3,000	Ocean color, sea surface temperature, vegetation
GMES-Sentinel-3/OLCI	2014	0.390–1.040 μm 21 bands	300–1200 m	12	2	1,250	Ocean color, sea surface temperature, sea surface topography
Landsat TM/ETM+	16 July 1982 (Landsat 4) 1 March 1984 (Landsat 5) 15 April 1999 (Landsat 7)	0.450–2.350 μm 7 bands (TM) 8 bands (ETM+)	30–120 m (TM) 15–60 m (ETM+)	8	16	185	Bathymetry, water turbidity, littoral processes, land cover and vegetation
SPOT/HRVIR	24 March 1998 (SPOT 4) 4 May 2002 (SPOT 5)	0.500–1.750 μm 5 bands	2.5–20 m	8	1–4	60	Bathymetry, littoral processes, land cover and vegetation, topography

Coastal Ecosystems, Table 1 (Continued)

Satellite/sensor	Date launched or planned	Spectral range and bands	Spatial resolution	Radiometric resolution (bit)	Revisit (day)	Swath width (km)	Specific coastal applications
IRS-1D/LISS	29 September 1997	0.520–1.700 μm 5 bands	5.2–23.5 m	8	24	70–148	Bathymetry, littoral processes, land cover and vegetation
Terra/ASTER	18 December 1999	0.520–11.65 μm 14 bands	15–90 m	8–12	16	60	Littoral processes, sea/land surface temperature, land cover and vegetation
IKONOS	24 September 1999	0.450–0.900 μm 5 bands	1–4 m	11	3–5	11	Bathymetry, water turbidity, littoral processes, land cover and vegetation, topography
QuickBird	18 October 2001	0.450–0.900 μm 5 bands	0.61–2.44 m	11	1–5	20–40	Bathymetry, water turbidity, littoral processes, land cover and vegetation, topography
OrbView-3	26 January 2003	0.450–0.900 μm 5 bands	1–4 m	11	<3	8	Bathymetry, water turbidity, littoral processes, land cover and vegetation, topography
GeoEye-1	6 September 2008	0.450–0.920 μm 5 bands	0.41–1.65 m	11	<3	15.2	Bathymetry, water turbidity, littoral processes, land cover and vegetation, topography
RADARSAT/SAR	4 November 1995 (RADARSAT-1) 14 December 2007 (RADARSAT-2)	C-band HH pol (RADARSAT-1) C-band HH, VV, VH, HV (RADARSAT-2)	6–100 m (RADARSAT-1) 3–100 m (RADARSAT-2)	16	1–4	20–500	Bathymetry, littoral processes, land cover and vegetation, topography
ENVISAT/ASAR	1 March 2002	C-band VV, HH, VV/HH, HV/HH, VH/VV	30 m	8	<3	50–100	Ocean currents and waves, land topography

been developed to measure energy patterns from different portions of the electromagnetic spectrum (see Jensen, 2007). Many such systems have been originally designed for terrestrial or open-ocean applications, among which some are relevant to coastal environments (Table 1). The following sections will provide an overview on remote sensing of coastal waters, submerged aquatic vegetation, benthic habitats, coastal wetlands, and watersheds. Note that for each application, acquiring accurate and consistent in situ data using ships, buoys, and field instruments is needed in the calibration and validation of remote signals and in algorithm development.

Remote sensing of coastal water quality indicators, such as chlorophyll-*a*, turbidity, dissolved organic matter (DOM), total nitrogen, temperature, and salinity, relies on the use of either a radiative transfer (see entry *Radiative Transfer, Theory*) algorithm or a statistical regression model. Airborne remote sensing systems are most useful for small, shallow, or optically complex coastal waters. Several existing spaceborne color scanners primarily designed for open oceans are of limited usefulness for nearshore coastal waters due to their relatively coarse spatial resolutions. Some researchers have been successful in remote sensing of nearshore coastal and estuarine waters by using satellite data from terrestrial sensors, such as Multispectral Scanner (MSS), Thematic Mapper (TM), Thematic Mapper Plus (ETM+), SPOT HRV, IKONOS, and QuickBird (see Yang, 2005b, 2008, 2009a, b). Compared to multispectral radiometers, hyperspectral sensors offer extremely high spectral resolution that can help discriminate complex bio-optical properties of coastal waters.

Mapping benthic habitats and submerged aquatic vegetation (SAV) can be accomplished through image interpretation or automated classification. The image interpretation approach is particularly suitable for aerial photography or high-resolution satellite imagery. It is based on the synthetic use of various image elements such as tone, pattern, texture, shape, size, and association in the identification and delineation. Individual benthic habitats or SAV polygons can be delineated manually or through an on-screen digitizer. Some field samples obtained through various instruments such as underwater video and still photography are needed to help develop image interpretation keys and verify the mapping accuracy. This image interpretation method can produce accurate results although it is quite labor intensive and highly dependent upon the skills and experience of an interpreter. On the other hand, there are various image classification methods, either supervised or unsupervised, which can be used for benthic habitat or SAV mapping. Image fusion from high-spectral- and high-spatial-resolution sensors can help improve the accuracy of benthic habitat or SAV mapping (Mishra, 2009). Change-detection analysis of benthic habitat or SAV can be made by using change vector analysis, principal component analysis, or map-to-map comparison.

Remote sensing has been used in wetland (including mangroves) mapping for several decades. Early inventories were largely based on the interpretation of aerial

photographs. Automated classification of satellite imagery facilitates the mapping of the spatial distribution of wetlands and the estimation of biomass or net productivity over a large area. Under unfavorable weather conditions, microwave or radar products can be used to help identify broad wetland classes. For species differentiation of wetland plants, broadband remote sensors may be problematic due to their inability to provide sufficient spectral details. Hyperspectral remote sensing allows the detection of subtle differences in canopy density, leaf and canopy structure, and biochemical properties, which can be further used to distinguish coastal wetland plant species. Additionally, LIDAR remote sensing can help estimate canopy structure of shrubs, marsh, grass, and other vegetation found in the littoral zone. The integration of spectral imagery and LIDAR data offer the potential to significantly improve the species classification and structural mapping of coastal plant communities (Nayegandhi and Brock, 2009).

Watershed landscape characterization affects coastal water quality by altering sediment, chemical loads, and hydrology, and therefore information on upstream landscape structure and patterns is indispensable for coastal ecosystem assessment. Remote sensing of coastal watershed landscape structure and patterns generally involves procedures of land cover map production and landscape metrics computation. Aerospace imagery from various terrestrial sensors, especially those equipped with a blue band, can be used to map land cover types in coastal watersheds. The production of an accurate land cover map for a coastal area is not a trivial task, mainly due to the presence of a variety of wetlands and vegetation covers, along with complex urban impervious materials and agricultural lands. Several strategies have been developed to improve automated land cover classification, which include GIS-based hierarchical classification and spatial reclassification, knowledge-based expert systems, artificial neural networks, fuzzy logic, and genetic algorithms. Landscape metrics can be derived from a land cover map, by which the landscape structure and pattern of a coastal watershed can be further assessed (Yang, 2009a).

Further research

This entry reviews several major coastal ecosystems and discusses how remote sensing can be used to characterize coastal waters, benthic habitats and submerged aquatic vegetation, coastal wetlands (including mangroves), and coastal watershed landscape characteristics.

While some significant progress has been made in remote sensing of coastal ecosystems, there are several major areas that deserve further research. Firstly, the current ocean color scanners are basically designed for offshore waters and are of limited usefulness for optically complex nearshore waters. Further research is needed to help design future ocean color radiometers appropriate for shallow coastal waters. Secondly, most of the algorithms for retrieving water quality measures were originally designed for open-ocean waters, and a significant

area for continuing research is the fundamental understanding of the functional linkage between water constituents and remote reflectance for coastal waters. Thirdly, more research is needed to advance the fundamental understanding of the relationship between the volumetric reflectance, the canopy density of SAV populations, water depth, and bottom reflectance parameters. This will help develop more realistic volumetric reflectance models, thus increasing the likelihood of accurate SAV mapping. Fourthly, there is an increased research demand to develop improved methods and technologies for resolving the spectral confusion between different land cover classes from medium-resolution imagery and for incorporating image spatial characteristics and ancillary data to improve land cover classification from high-resolution imagery. Fifthly, more research is needed to advance the fundamental understanding of the relationship between landscape patterns and ecological processes. Lastly, continuing research efforts are needed to help acquire good and sufficient in situ data for building comprehensive spectral libraries of different coastal plant species and for calibrating remote signals and verifying information extraction algorithms for coastal environments.

Bibliography

- Beatley, T., Brower, D. J., and Schwab, A. K., 2002. *An Introduction to Coastal Zone Management*, 2nd edn. Washington, DC: Island Press, p. 329.
- Burke, L. A., Kura, Y., Kassem, K., Revenga, C., Spalding, M., and McAllister, D., 2001. *Pilot Analysis of Global Ecosystems: Coastal Ecosystems*. Washington, DC: World Resources Institute, p. 77.
- Council, N. R., 2000. *Clean Coastal Waters: Understanding and Reducing the Effects of Nutrient Pollution*. Washington, DC: National Academy Press, p. 200.
- Hinrichsen, D., 1998. *Coastal Waters of the World: Trends, Threads, and Strategies*. Washington, DC: Island Press, p. 298.
- Hobbie, J. E., 2000. *Estuarine Science: A Synthetic Approach to Research and Practice*. Washington, DC: Island Press, p. 539.
- Jensen, J. R., 2007. *Remote Sensing of the Environment: An Earth Resource Perspective*. Upper Saddle River, NJ: Prentice Hall, p. 592.
- Klemas, V., 2009. Sensors and techniques for observing coastal ecosystems. In Yang, X. (ed.), *Remote Sensing and Geospatial Technologies for Coastal Ecosystem Assessment and Management*. Berlin/Heidelberg: Springer, pp. 17–43.
- Mann, K. H., 2000. *Ecology of Coastal Waters with Implications for Management*, 2nd edn. Malden, Massachusetts: Blackwell Science, p. 406.
- Mishra, D. R., 2009. High-resolution ocean color remote sensing of coral reefs and associated benthic habitats. In Yang, X. (ed.), *Remote Sensing and Geospatial Technologies for Coastal Ecosystem Assessment and Management*. Berlin/Heidelberg: Springer, pp. 171–210.
- Nayegandhi, A., and Brock, J. C., 2009. Assessment of coastal-vegetation habitats using airborne laser remote sensing. In Yang, X. (ed.), *Remote Sensing and Geospatial Technologies for Coastal Ecosystem Assessment and Management*. Berlin/Heidelberg: Springer, pp. 365–390.
- Selman, M., Greenhalgh, S., Diaz, R., and Sugg, Z., 2008. *Eutrophication and Hypoxia in Coastal Areas: A Global Assessment of the State of Knowledge*. Washington, DC: World Resources Institute, p. 6.
- Yang, X., 2005a. Remote sensing and GIS applications for estuarine ecosystem analysis: an overview. *International Journal of Remote Sensing*, **26**(23), 5347–5356.
- Yang, X., 2005b. Special issue: remote sensing and GIS for estuarine and coastal ecosystem analysis. *International Journal of Remote Sensing*, **26**(23), 5163–5356.
- Yang, X., 2008. Theme issue: remote sensing of the coastal ecosystems. *ISPRS Journal of Photogrammetry and Remote Sensing*, **63**(5), 485–590.
- Yang, X., 2009a. Integrating satellite imagery and geospatial technologies for coastal landscape pattern characterization. In Yang, X. (ed.), *Remote Sensing and Geospatial Technologies for Coastal Ecosystem Assessment and Management*. Berlin/Heidelberg: Springer, pp. 461–494.
- Yang, X., 2009b. *Remote Sensing and Geospatial Technologies for Coastal Ecosystem Assessment and Management*. Berlin/Heidelberg: Springer, p. 561.
- Yang, X., Damen, M. C. J., and van Zuidam, R., 1999. Use of Thematic Mapper imagery and a geographic information system for geomorphologic mapping in a large deltaic lowland environment. *International Journal of Remote Sensing*, **20**(4), 568–591.

Cross-references

[Data Processing, SAR Sensors](#)
[Fisheries](#)
[Forestry](#)
[Geophysical Retrieval, Overview](#)
[Global Land Observing System](#)
[Lidar Systems](#)
[Ocean Data Telemetry](#)
[Ocean Measurements and Applications, Ocean Color](#)
[Ocean Modeling and Data Assimilation](#)
[Ocean Surface Topography](#)
[Pattern Recognition and Classification](#)
[Radiative Transfer, Theory](#)
[Sea Level Rise](#)
[Sea Surface Salinity](#)
[Sea Surface Temperature](#)
[Vegetation Indices](#)
[Water Resources](#)
[Wetlands](#)

COMMERCIAL REMOTE SENSING

William Gail
 Global Weather Corporation, Boulder, CO, USA

Definition

Commercial. Pertaining to organizations or activities associated with buying and selling products and services.

Introduction

Commercial remote sensing is remote sensing that produces information intended for sale on open commercial markets. It is generally performed by companies or through public-private partnerships rather than by governments. The commercial market includes remote sensing performed primarily from aerial and spaceborne

platforms. Due to the strategic nature of remote sensing, commercial interests are nearly always intertwined with national interests, at least for spaceborne systems. Governmental policies authorizing and regulating commercial remote sensing generally reflect objectives for national security, technology development, international relations, and economic positioning. The history of the industry, and its current configuration, reflect this tension between economic and national interests. Demand for commercial remotely sensed data has increased in recent years, in part due to the widespread availability of geographic information systems (GIS) that can display and analyze the imagery and online means for distributing the results.

History

Aerial commercial remote sensing has been in existence for nearly a century. A significant focus of the industry is photogrammetry, the making of geometrically accurate high-resolution image mosaics covering large areas such as cities and states. Today, such mosaics are routinely produced with image resolutions as good as 15 cm and sometimes better. Until recently, the mosaics were generally panchromatic, but color imagery has become common with the introduction of digital sensors during the 2000s. In addition to optical photogrammetry, commercial remote sensing includes work with synthetic aperture radar (SAR), hyperspectral imagers, and LIDAR sensors. Both SAR and LIDAR are commonly used for measuring topography and related three-dimensional imaging. The aerial remote sensing industry consists of both commercial and government-owned entities, often in competition.

The roots of spaceborne commercial remote sensing extend back to the early 1970s with efforts by the USA, France, and India to provide commercial adjuncts to their national Earth monitoring systems. These early efforts laid the groundwork for today's industry with several dozen commercial systems in operation and a global market that is multibillion dollars (Modello et al, 2004).

In 1972, the USA launched the Earth Resources Satellite 1, which was subsequently renamed Landsat. In part a Cold War effort to demonstrate US leadership in Earth resources monitoring, its commercial value became recognized soon after launch. By the early 1980s, with the various interested parties all seeking a better way to access Landsat's imagery (National Research Council, 1985), a strong effort was made to change the regulatory environment governing Landsat. The result was the Land Remote Sensing Commercialization Act of 1984, which for the first time established the basis for commercial remote sensing operations authorized by the USA. As a result, in 1985 Landsat was partially privatized through a partnership with Earth Observation Satellite Corporation (EOSAT), itself a partnership of Hughes Aircraft and RCA. EOSAT was given a contract to operate the system for 10 years and build follow-on satellites, obtaining rights for commercial sale of the data.

Around the same time, the French *Système pour l'Observation de la Terre* (SPOT) satellite series was being developed. To encourage commercialization, a public-private partnership called Spot Image was created by the French government in 1986. They have since launched a series of increasingly capable satellites, starting with the 10 m resolution SPOT-1 in 1986. In India, the Indian Space Research Organization (ISRO) initiated a commercially oriented Earth-observing program following the success of their early demonstration satellites Bhaskara-1/2 launched in 1979 and 1981. The Indian Remote Sensing (IRS) series of satellites was initiated with the IRS-1 launch in 1988 and continues as a robust program today.

As demand grew and technology improved, these early efforts spurred further commercial interest beginning in the 1990s. The first truly commercial (not public-private partnerships) remote sensing satellite operations were plagued by early failures. EarlyBird was launched by EarthWatch (originally called Worldview and changed in 2001 to DigitalGlobe) in 1997 but functioned for only 3 days. Space Imaging launched their first satellite Ikonos in 1999 but it failed to reach orbit. Space Imaging succeeded with their second Ikonos satellite later in 1999. EarthWatch's second satellite, QuickBird, also suffered a launch failure in 2000, but the third satellite Quickbird-2 succeeded in 2001. ORBIMAGE joined the group in 2003 with the launch of Orbview-3, but only after Orbview-4 was lost on launch. By the early 2000s, these firms had become increasingly dependent on contracts from the US National Geospatial-Intelligence Agency (NGA) for revenue. In 2005, ORBIMAGE acquired Space Imaging for \$58 million after it failed to win the latest rounds of NGA contracts. ORBIMAGE subsequently renamed itself GeoEye.

More recently, the intelligence community has been a significant partner in the development of commercial remote sensing capability. In the USA, large data purchase contracts such as NextView, ClearView, and EnhancedView by NGA have provided anchor funding for commercial remote sensing companies. NextView contracts, for example, are service-level agreements that specify requirements for imagery purchases from satellites still to be developed. They are designed to provide a long-term government partnership commitment that enables the private sector partner to finance and develop new satellites that meet the requirements. To motivate a commercial market, the agreements are nonexclusive, meaning that the companies can sell the same imagery to other customers. A 2007 independent report chartered by the US National Reconnaissance Office (Marino, 2007) defined various approaches for intelligence community use of commercial remote sensing along with the benefits and risks of those approaches.

The Landsat Data Continuity Mission, initiated in the early 2000s as the follow-on to Landsat-7, was originally envisioned as a strong public-private partnership but has since evolved into a more traditional system acquisition with the government retaining responsibility for data

processing and distribution. NASA also used the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) mission to promote commercialization. SeaWiFS was built for NASA under partnership with Orbital Sciences Corporation and launched in 1997. ORBIMAGE (renamed GeoEye), a spin-off from Orbital Sciences, was given the responsibility for operating the satellite (renamed Orbview-2 for commercial purposes) and for selling the data, with NASA itself purchasing data.

The mid-2000s saw the entrance of many new commercial players with innovative approaches to the market. RapidEye, a public-private partnership, developed and launched a series of five satellites designed to provide rapid-refresh multispectral optical imagery. The TerraSAR satellite system owned and operated by InfoTerra, another public-private partnership, focused on SAR imagery. Surrey Satellite Technology Ltd. (SSTL) developed a multi-government collaboration that built and launched a series of small satellites called the Disaster Monitoring Constellation (DMC), each owned and controlled by a different government (Algeria, Nigeria, Turkey, Britain, China, Spain) but operated jointly through a wholly owned subsidiary of SSTL called DMC International Imaging.

Despite today's successes, a number of companies with ambitious commercial business plans have attempted to enter the market and failed. Resource21 was focused on providing short-revisit multispectral imagery for use in agriculture starting in the 1990s. Their plans for a satellite system focused on this market did not proceed. Around the same time, AstroVision began pursuing a new market for imagery taken from geostationary orbit but had not been able to develop the system after more than a decade. In 2011, RapidEye filed for the German equivalent of bankruptcy protection. Even after several decades of market development, spaceborne commercial remote sensing remains a risky business with considerable investment and high likelihood of failure (National Research Council, 2002; Group on Earth Observations, 2005).

Technology

Of the many spaceborne remote sensing technologies developed over the years for scientific and defense purposes, those that have been commercialized are largely constrained to high-resolution imagery. Initially, this meant panchromatic optical imagers, although inclusion of color bands quickly became routine. One current trend is an increase in the number of color bands and the extension of these bands to infrared wavelengths. Hyperspectral imagers represent the next logical step in this progression, although no company has yet announced plans for doing so. Canada pioneered use of synthetic aperture radar in the commercial realm, with the 1995 launch of Radarsat, and a number of commercial radars are operating now. A significant technology advance occurred with the reduction in size and weight of spacecraft platforms, allowing the introduction of low-cost multi-satellite

constellations such as DMC and RapidEye. A number of companies have built businesses around the sale of satellite systems and technology to nations that do not have the capability to build their own.

Aerial remote sensing technologies have followed a somewhat different path. Because aerial systems allow access to the sensor itself, film remained a viable sensing medium far longer than with spaceborne systems. In the early 2000s, highly capable digital aerial cameras were introduced, and by 2010 most of the industry had transitioned to digital imagery. Related technologies, such as widespread availability of GPS positioning and accurate inertial sensors, made digital sensors even more effective.

Regulation

Commercial remote sensing is regulated on a nation-by-nation basis, involving both policy and law. In general, the purpose of such regulation is to make data available for legitimate scientific and commercial uses while at the same time protecting national security interests (Gabrynowicz, 2007). A critical aspect of such regulation is the use of licensing that authorizes companies to operate commercial remote sensing systems and defines the constraints on that operation. Central to this are rules for ownership, dissemination, and use of data, particularly with regard to resolution and timeliness of distribution.

The most developed regulatory framework is that of the USA. The earliest formal legislation was the *Land Remote Sensing Commercialization Act of 1984*, a by-product of the Reagan-era policies to promote privatization of government assets. The Act was driven by the specific need to create better access to Landsat imagery, but it also introduced a more general framework for licensing and regulation of commercial remote sensing systems. This was updated through the *Land Remote Sensing Policy Act of 1992*, which reversed the commercialization of Landsat contained in the 1984 Act, but provided further legislation promoting private sector remote sensing.

Despite these legislative acts, uncertainty remained regarding how licensing of commercial remote sensing systems would actually occur. This uncertainty was initially resolved in 1994 through a presidential policy statement known as the *US Policy on Foreign Access to Remote Sensing Space Capabilities* (PDD-23). It stated clearly the dual-purpose policy goal: "the fundamental goal of our policy is to support and to enhance US industrial competitiveness in the field of remote sensing space capabilities while at the same time protecting US national security and foreign policy interests." PDD-23 provided the administrative mechanisms needed to make the 1992 legislative act effective, including four fundamental principles: (a) the presumption that US licenses would only be issued for systems with capability no better than what is commercially available from other suppliers in the world marketplace, (b) tight control of technology exports in the form of satellite systems and components,

(c) management of system capabilities (rather than data distribution) as a means for constraining imagery access by undesirable parties, and (d) “shutter control” that allows the government to stop all commercial remote sensing when appropriate national interests are identified.

Licenses are further constrained by an amendment to the 1997 Defense Authorization Act (known as the Kyl-Bingaman amendment) that specifically limits imagery taken over Israel to no better than that “available from commercial sources.” Throughout much of the 1990s, for example, this limit was set to 2 m. The *Commercial Remote Sensing Policy of 2003* (NSPD-27), which superseded PDD-23, did not change these licensing provisions but strengthened the commercial sector by encouraging drawing on commercial capabilities to the “maximum practical extent” when serving government needs.

Under this framework, authority for licensing commercial remote sensing is allocated to the Department of Commerce. Within the Department of Commerce, the NOAA Office of Space Commercialization provides policy guidance while licenses themselves are issued by the NOAA Commercial Remote Sensing Regulatory Affairs Office. Between 1993 and 2000, 17 licenses were issued. In 2008, a total of 19 licenses were active covering 45 satellites with 10 launched. This office is advised in the effort by an independent board called the Advisory Committee on Commercial Remote Sensing (ACCRES). NOAA is responsible for assessing the state of the industry and determining, for example, the resolution of systems that can be licensed. Among the provisions included originally in licenses is that no imagery can be distributed for at least 24 h after it is collected, a constraint designed to minimize use of US-collected imagery by enemies for targeting US military positions. After lengthy interagency governmental deliberation, this provision was dropped in 2007.

Canada also has a relatively mature and transparent regulatory framework built around the *Canadian Space Agency Act of 1990* and the *Remote Sensing Space Systems Act of 2005*. The USA and Canada have a formal agreement concerning operation of commercial remote sensing systems. India, on the other hand, has no relevant law but rather a comprehensive set of policies (Rao et al., 2002; Gabrynowicz, 2007). In Germany, regulation is provided through the *2008 Satellite Data Security Law*, which specifies that approval for commercial remote sensing activities is applied not to each satellite development but rather to each data request on a case-by-case basis. Japan established a *Basic Space Law* in 2008 and released a *Basic Plan for Space Policy* based on that law in 2009. Other nations tend to have less open or less specific regulatory guidelines or to regulate on a satellite-by-satellite basis (Gabrynowicz, 2007). Some govern commercial remote sensing through general space policies, while others have only informal policies. Only a limited amount of international agreement is available to guide commercial remote sensing. In 1986, the United Nations released

a set of principles concerning remote sensing from space (United Nations, 1986), reflecting the desire for remote sensing activities to be used for the benefit of all nations.

Current examples

Three recent trends have had enormous impacts on aerial remote sensing. The first is the transition from analog to digital technologies within optical imagers, as discussed previously. Alternate sensing technologies, including radar and LIDAR, are still establishing their niches within overall commercial offerings. The second is advances in computers and software, making it feasible to work with large imagery volumes in professional domains such as GIS and photogrammetry. The third is end-user applications, such as online mapping, that make such imagery and related value-added products accessible to consumers. Today, there is a robust commercial industry dedicated to the collection of aerial remote sensing and related industries for value-added products, software, and instrument development. Despite these technology changes, the industry business models have evolved more slowly and remain characterized by smaller companies. The nature of the industry varies widely by country, with some less robust due to strong competing government capabilities.

For the foreseeable future, high-resolution electro-optical imagery will remain the central focus of spaceborne commercial remote sensing. At the present time, the USA retains leadership in the area of commercial electro-optical imagery, but other nations lead in the area of radar imagery. Considerable innovation is being applied to introduce capabilities such as rapid revisit and advanced multispectral imagery that serve customers beyond those needing the highest resolution.

GeoEye’s latest satellite GeoEye-1 was launched in 2008 and was developed with funding raised in part through an NGA NextView contract. GeoEye-1 provides panchromatic imaging with 0.41 m resolution (although the operating license requires this to be resampled to no better than 0.5 m before distribution) and multispectral imagery with resolution 1.7 m. A second GeoEye satellite is expected sometime by 2013. DigitalGlobe’s Worldview-1 (launched 2007), also funded in part through an NGA contract, produces 0.5 m panchromatic imagery. Worldview-2 (launched 2009) provides 0.46 m panchromatic imagery and 8-band color with 1.8 m resolution. DigitalGlobe also operates the older QuickBird satellite.

Spot Image operates several SPOT satellites and began launching the more advanced Pleiades series beginning in 2011. The Pleiades satellites provide panchromatic imagery at 0.7 m resolution and 4-band color at 2 m resolution. Spot Image also has distribution rights to other national systems, including Korea’s KOMPSAT and Republic of China’s FORMOSAT. Russia launched the Resurs-DK1 satellite, with 0.9 m resolution, in 2006 and markets data through SOVZOND JSC. Russia has also leveraged imagery from older surveillance satellites, digitizing that

imagery and selling it through commercial channels. The Israeli EROS-A/B imagery satellites, launched between 2000 and 2006 with resolution as good as 0.7 m, are owned and operated by an Israeli-founded international company ImageSat International.

The Indian Remote Sensing (IRS) series of satellites is the largest single commercially oriented remote sensing system in the world. It now includes three satellite classes focused on land resources, ocean resources, and cartography. ResourceSat-1/2 satellites, launched in 2003 and 2009, have 6 m pan resolution and 3-band color at the same resolution. OceanSat-2 was launched in 2009 and focused on ocean color imaging, carrying an 8-band multispectral imaging and Ku-band scatterometer. The Cartosat series has had four launches starting in 2005, with pan resolution better than 1 m. Commercial image distribution is the responsibility of ISRO's commercial entity Antrix.

Within eastern Asia, neither China nor Japan has commercial remote sensing capability, though each has a robust remote sensing program. The Korean KOMPSAT-1/2 satellites, launched in 1999 and 2006 by the Korean Aerospace Research Institute (KARI), have resolution of 4-band color with 4 m resolution. The Republic of China FORMOSAT-1/2 series, launched in 1999 and 2004, have a resolution of as good as 1 m and 4-band color with 8 m resolution. The imagery from both KOMPSAT and FORMOSAT is currently marketed by Spot Image.

Several companies have pursued somewhat different approaches to the commercial market, focusing on improved repeat cycles rather than highest resolution. Surrey Satellite Technology Ltd. (SSTL) developed a multi-government collaboration that built and launched a series of small satellites called the Disaster Monitoring Constellation (DMC), each owned and controlled by a different government (Algeria, Nigeria, Turkey, Britain, China, Spain) and operated jointly through a wholly owned subsidiary of SSTL called DMC International Imaging. The German RapidEye satellite constellation, a five-satellite system designed for rapid-refresh multispectral imagery, was developed as a public-private partnership with a diverse set of public and private investors largely from Germany and Canada. The satellites were launched on a single launch vehicle in 2008.

Radar imagery has proven to be of strong commercial interest. Radarsat International was formed in 1989 by a consortium of Canadian companies with the purpose of processing and distributing the data from Radarsat-1, launched in 1995. One of the shareholders, Macdonald Dettwiler and Associates, bought out the others in 1999 and formed a deeper partnership with the Canadian government to build and operate Radarsat-2, launched in 2007. The Radarsat-2 resolution of 3 m substantially improved on the 8 m resolution of Radarsat-1. TerraSAR-X, launched in 2007, is a public-private partnership between EADS (through their InfoTerra company) and the German Aerospace Center (DLR). It produces radar imagery with resolution up to 1 m including a variety of multipolarization and interferometric data

products. A companion satellite, TanDEM-X, was launched in 2010, allowing the system to operate in interferometric mode so as to perform high-resolution topographic mapping. COSMO-SkyMed is a dual-use system of four satellites with 1 m resolution and polarimetric/interferometric capability funded by the Italian Space Agency, the Italian Ministry of Defense, and the commercial company Telespazio. The first launch was in 2007 and the last in 2010.

Summary

The commercial market today is served by both aerial and spaceborne sectors, with some overlap. Over the last decade, the relative roles of each sector have become more clearly defined within the market. The aerial sector continues to be dominated by smaller companies and governmental entities, with significant national and even regional variability. The spaceborne sector is moving rapidly from an early exploratory phase, in which new technologies, business models, and regulatory environments were being assessed, to a mature phase with well-established success paths. The entire space industry is becoming more commercialized, from launch vehicles to spaceports, establishing a positive context within which remote sensing's own commercialization takes place. The remote sensing industry as a whole is moving from a platform-centric view to a services-centric view in which the features of the data are more important than the features of the sensors. Governments are becoming more focused on remote sensing as an important element of their space programs. They are migrating to smaller but more capable satellites, and they view commercialization or dual use as an important element of their efforts. Capabilities of commercial systems are increasingly diverse, from the growing use of radar to the selection of optical bands to the availability of short-revisit imagery. All trends suggest that the importance of commercial remote sensing will continue to grow in the coming years.

Bibliography

- Council, N. R., 1985. *Remote Sensing of the Earth from Space: A Program in Crisis*. Washington, DC: The National Academies.
- Council, N. R., 2002. *Toward New Partnerships in Remote Sensing: Government, the Private Sector, and Earth Science Research*. Washington, DC: The National Academies.
- Gabrynowicz, J., 2007. *The Land Remote Sensing Laws and Policies of National Governments: A Global Survey*. Report for U.S. Department of Commerce/National Oceanic and Atmospheric Administration's Satellite and Information Service Commercial Remote Sensing Licensing Program.
- Group on Earth Observations, 2005. *Global Earth Observations System of Systems GEOSS: 10-Year Implementation Plan Reference Document*. Noordwijk: ESA Publications Division.
- Marino, P., 2007. *Independent Study of the Roles of Commercial Remote Sensing in the Future National System for Geospatial-Intelligence (NSG)*. Falls Church, VA: Defense Group.
- Modello, C., Hepner, G. F., and Williamson, R. A., 2004. *10-Year Industry Forecast, Phase I-III - Study Documentation*. Bethesda, MD: ASPRS.
- Rao, M., Jayaraman, V., and Sridharamurthi, K. R., 2002. Issues for a remote sensing policy and perspective of the Indian Remote

Sensing Programme. In *Proceedings from The First International Conference on the State of Remote Sensing Law*, pp. 47–61.
United Nations, 1986. *Principles Relating to Remote Sensing of the Earth from Outer Space*. New York: United Nations.

Cross-references

Cost Benefit Assessment
Emerging Applications
Observational Systems, Satellite
Policies and Economics
Public-Private Partnerships
Radar, Synthetic Aperture

COSMIC-RAY HYDROMETEOROLOGY

Darin Desilets¹ and Marek Zreda²

¹Hydroinnova LLC, Albuquerque, NM, USA

²Department of Hydrology and Water Resources,
University of Arizona, Tucson, AZ, USA

Definitions

Cosmic-ray hydrometeorology. The science of measuring hydrologic variables through their effects on secondary cosmic-ray intensity.

Primary cosmic ray. A charged particle, usually a proton, traveling toward Earth at relativistic speed.

Secondary cosmic ray. An energetic proton, neutron, or other subatomic particle generated as a consequence of primary cosmic rays colliding with the Earth.

Introduction

Cosmic rays continually bombard Earth, giving rise to a small but measureable flux of background neutrons at the land surface. These ambient neutrons respond strongly to the presence of land surface water in the form of soil moisture and snow or more specifically to the hydrogen which that water contains. The unique ability of hydrogen to influence neutron intensity has been known since the discovery of the neutron itself in the 1930s, when the mysterious nonionizing radiation was first identified through its ability to scatter hydrogen nuclei from paraffin (Chadwick, 1932), losing substantial fraction of its energy in the process. By the late 1950s, soil scientists began applying neutron scattering principles to the field determination of soil water content by lowering radioisotopic neutron sources and colocated neutron detectors down bore holes and measuring the intensity of backscattered neutrons (Gardner and Kirkham, 1952). It was later demonstrated that water content in the shallow subsurface and snow water equivalent depth could be obtained passively by measuring background cosmic-ray neutrons with detectors buried in the top meter of soil (Kodama et al., 1985), although the method was never widely adopted. More recently, Zreda et al. (2008) showed the feasibility of noninvasively measuring soil water content through

aboveground measurements of cosmic-ray neutron intensity. This technique operates at a scale of tens of hectares, which fills an important gap between the scales of invasive point measurements and satellite remote sensing footprint.

Neutron interactions in soil

The transmission of fast neutrons through bulk matter is profoundly influenced by the presence of hydrogen, which at the land surface is present mainly in the form of liquid and solid water and plant carbohydrates. Hydrogen is uniquely effective in moderating (slowing) neutrons by virtue of its low mass and relatively large elastic scattering cross section, which is a measure of the probability of interacting elastically with a neutron. As with any two particles having the same mass, a fast neutron can theoretically be brought to rest through a single head-on collision with hydrogen (Glasstone and Edlund, 1952). The fewer collisions needed to moderate a fast neutron, the lower the fast neutron intensity will be. Elastic collisions with hydrogen and other light nuclei progressively moderate a fast neutron until it is either absorbed by a nucleus or is reduced to a velocity on the order of the thermal motions of surrounding molecules, at which point there is no net change in energy through subsequent collisions. A distinguishing characteristic of thermal neutrons is their strong tendency to be absorbed by nuclei. Common absorbing elements in the soil matrix include major elements such as Fe, Ca, K, and trace elements with unusually high absorption cross sections, such as B, Gd, and Sm.

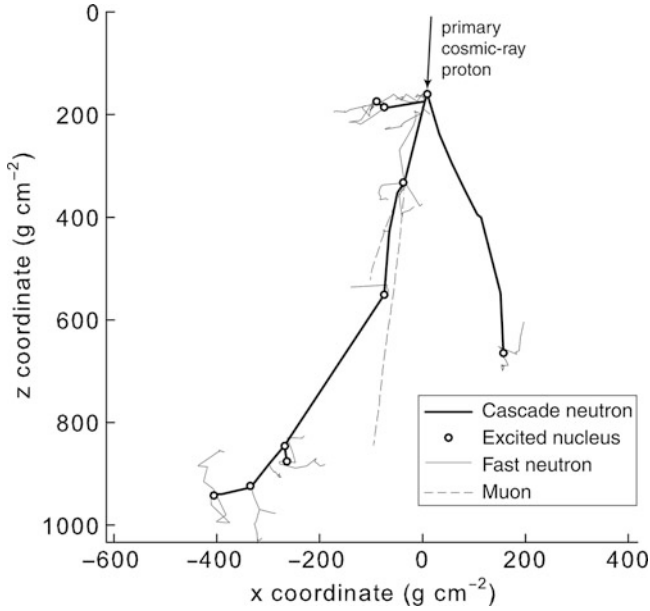
Cosmic-ray neutrons

Cosmic-ray neutrons are an ever-present part of the land surface radiation environment. They are a by-product of chain reactions initiated at the top of the atmosphere by primary cosmic rays (Simpson, 1951) (A simulated particle cascade created by a 10 GeV primary interacting with nitrogen is shown in Figure 1). The primary radiation is composed of highly energetic particles, mainly protons and helium nuclei, which are believed to have been accelerated in shock waves associated with supernovas occurring throughout the Milky Way (Uchiyama et al., 2007). Energetic primaries collide with atmospheric gas molecules, unleashing cascades of secondary protons, neutrons, and other subatomic particles, some of which penetrate to sea level.

The neutrons utilized for passive water content measurements are generated mainly by cascade neutrons interacting with matter. Fast neutrons are produced in two types of interactions. A cascade neutron can transfer kinetic energy to an entire target nucleus, raising it to an excited energy state. The nucleus then cools off by “evaporation,” that is, the emission of fast neutrons in random directions. Cascade neutrons with higher energies will tend to interact at the surface of a nucleus, dislodging the outermost neutrons in a mostly forward direction (Krane, 1987). Regardless of how they are produced, fast neutrons

are scattered elastically in random directions until they are eventually absorbed by soil or atmospheric nuclei.

Fast neutron intensity at the land surface reflects the equilibrium between production, moderation, and absorption of neutrons in the ground and atmosphere (Figure 2).



Cosmic-Ray Hydrometeorology, Figure 1 Atmospheric particle cascade simulated with the radiation transport code Monte Carlo N-Particle eXtended (MCNPX – Pelowitz, 2005). A 10 GeV primary cosmic-ray proton collides with atmospheric nitrogen, triggering a particle cascade that reaches sea level. Fast neutrons are generated at each collision marked by a circle. The fast neutrons are scattered in random directions as they are moderated and eventually become captured.

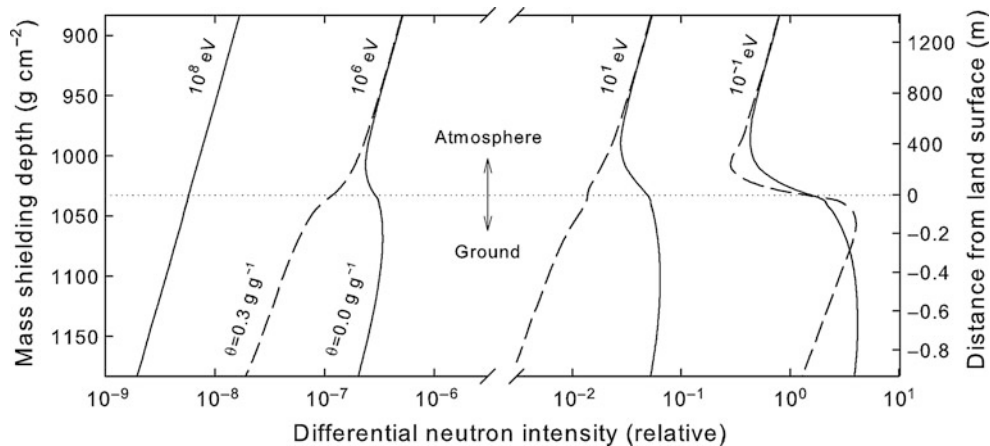
Neutrons are scattered between the ground, which tends to be the better moderator when wet, and the atmosphere, which is the better thermal neutron absorber. Any change in water content at the land surface disturbs this equilibrium. An increase in the amount of soil water or snow decreases the intensity in the fast to epithermal region because neutrons are more efficiently reduced to lower energies through collisions with hydrogen. Conversely, thermal neutron intensity first increases with increasing water content and then decreases monotonically. This behavior is explained by the competing roles of hydrogen as an absorber and moderator. Initially, a small increase in water content rapidly increases the rate of thermalization, which increases the thermal neutron flux. But above a few percent gravimetric soil moisture content, the role of hydrogen as a moderator is challenged by its tendency to absorb neutrons.

Retrieval of soil moisture

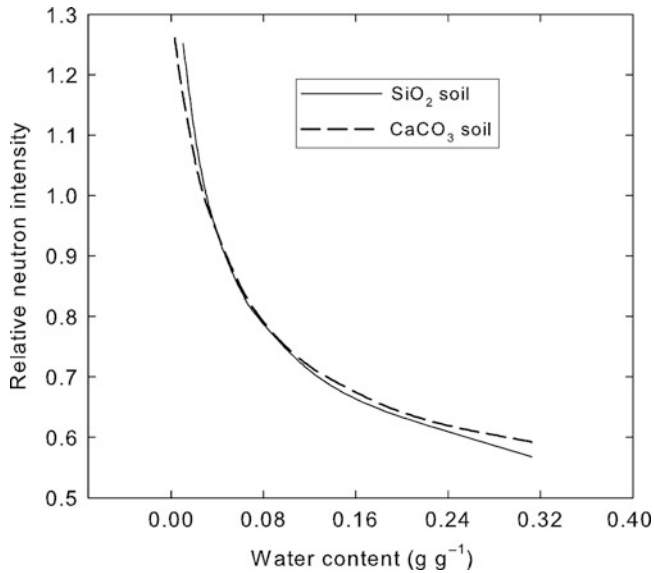
For a wide range of soil compositions, a “universal” shape-defining function can be used to convert neutron counting rates to soil water content for typical silica-dominated soils (Figure 3). This function is valid for neutrons in the epithermal to fast part of the spectrum (10^0 – 10^6 eV), where neutron absorption is minor. A calibration curve for soil water content has been obtained by fitting simulated ground-level neutron fluxes to the semiempirical shape-defining equation:

$$\theta(\phi/\phi_r) = \frac{1}{B} \left(\frac{A}{(\phi/\phi_r) - D} - C \right) \quad (1)$$

where (ϕ/ϕ_r) is the neutron intensity normalized to a reference soil moisture state. The dependence of land surface neutron intensity in the epithermal to fast energy range on gravimetric soil water content is represented in



Cosmic-Ray Hydrometeorology, Figure 2 Depth profiles of neutron intensity near the land surface for different energies simulated with MCNPX. Between 10^1 and 10^6 eV, where elastic scattering interactions dominate, the shape of the profile and its sensitivity to soil moisture are remarkably constant. At lower energies ($<10^{-1}$ eV), thermal neutron absorption becomes important and the shape of the profile reflects the strong contrast between the absorbing properties of the ground and atmosphere. At higher energies ($>10^8$ eV) the profile reflects exponential attenuation expected for cascade neutrons.



Cosmic-Ray Hydrometeorology, Figure 3 The dependence of neutron intensity on soil water content at 0–2 m above the ground according to MCNPX calculations for a soil matrix with pure SiO_2 . For comparison, a curve for a pure CaCO_3 soil matrix is also shown.

silica-dominated soils by $A = 0.562$, $B = 0.060$, $C = 0.942$, $D = 0.449$, with the reference state dry soil. The shape of the function is not significantly altered by differences in soil texture, salinity, bulk density, or moderate amounts of carbonates or organic matter in soil.

Although the shape of the calibration function is largely invariant, the absolute “source” neutron intensity is highly variable across the surface of the Earth. In other words, the calibration function can translate on the intensity axis depending on location. Elevation differences are responsible for the biggest differences in source intensity. The elevation effect is more accurately described as a function of the mass shielding depth at a site, which is the product of atmospheric depth and air density and is expressed in units of g cm^{-2} . Local barometric pressure readings are usually an acceptable proxy for mass shielding depth. Neutron-generating cascades are attenuated exponentially as a function of mass shielding according to

$$\phi_2 = \phi_1 \exp[(x_1 - x_2)/\Lambda] \quad (2)$$

where ϕ_1 and ϕ_2 are the neutron intensities at depths x_1 and x_2 (g cm^{-2}) and the attenuation length Λ is $\sim 130 \text{ g cm}^{-2}$ at high to mid-latitudes (Desilets et al., 2006). According to this relationship, the neutron intensity at an elevation of 3,000 m (750 g cm^{-2}) is almost nine times greater than at sea level ($1,033 \text{ g cm}^{-2}$). Neutron intensity decreases by about half from high and mid-latitudes to the equator due to stronger magnetic shielding of primary cosmic rays at lower geomagnetic latitude. The elemental composition of soil may also have some effect on neutron source intensity because of differences

between elements in the number of neutrons emitted following excitation.

The spatial variability in cosmic-ray intensity means that the calibration function must be normalized to the local neutron source strength. This can be accomplished by obtaining at least one field calibration point. Because the radius of influence is large, many field samples are usually needed in order to obtain an areally representative average moisture (e.g., Western and Blöschl, 1999; Famiglietti et al., 2008). Different strategies to obtaining this point may be employed. One consideration is that average soil water content is ideally in the middle of the anticipated moisture range. Another consideration is that samples should be collected when the distribution of soil moisture is expected to be fairly uniform, in order to reduce the number of samples required for representativeness. A calibration is transferrable to another site if the topography, biomass concentration, and soil composition are similar between the two sites and elevations and latitudes are similar. Differences in source neutron intensity related to elevation and latitude can be compensated for by applying published scaling factors for neutron intensity (e.g., Desilets et al., 2006).

Counting rates should also be corrected for fluctuations in neutron source intensity over time. These are related mainly to variations in barometric pressure and solar activity. Corrections for barometric pressure can be made with Equation 2 using local pressure data and the local attenuation length for neutron-generating cosmic rays. Changes related to solar activity can be corrected using publicly available data from the global network of neutron monitors (Kuwabara et al., 2006).

Measuring neutron intensity

Neutron detectors tend to be most sensitive over a limited range of energies. The optimal sensitivity for soil moisture measurements is in the epithermal (10^0 – 10^1 eV) range because a reasonably high count rate can be achieved while sensitivity to neutron absorbers is minimized. Although the shape of the calibration function is nearly constant up to 10^6 eV, neutron intensity drops off rapidly with energy according to a $1/E$ law (Glasstone and Edlund, 1952). Measurements at higher energies therefore require larger or more efficient detectors or longer averaging times in order to compensate for lower neutron intensity.

The analytical precision of soil moisture determinations is governed by Poissonian statistics, which assumes that neutron counts are uncorrelated. The coefficient of variation is given by $N^{-0.5}$, where N is the counting rate. A precision of better than 2–3 % for 1 h of counting is easily achieved at sea level using portable equipment.

Sample volume

A major advantage of subaerial cosmic-ray measurements is that a large area can be sampled noninvasively from the ground or a low-flying aircraft. The radius of influence for 86 % (two e-fold drops) of the counts is 350 m at sea level

for a ground based, omnidirectionally sensitive neutron detector. Several factors are responsible for the large radius of influence: the neutron source is distributed across the land surface, the mean free path for atmospheric collisions is on the order of 30 m, and trajectories are randomized through collisions in the atmosphere. The footprint increases with elevation in proportion to the atmospheric collision mean free path length, which is inversely proportional to atmospheric pressure. The measurement depth depends strongly on soil moisture, ranging from 0.6 m in dry soil to 0.2 m in saturated soil for 86 % of the response.

Summary

Soil water content can be inferred from subaerial measurements of cosmic-ray neutron intensity. The hydrogen in soil water dominates the moderating power in the land surface environment. Through its ability to moderate and absorb neutrons, hydrogen exerts a strong control on neutron fluxes in the fast to thermal energy range. Cosmic-ray measurements are passive, noninvasive, noncontact, and represent a sample area of tens of hectares and a depth of tens of centimeters. The method has moderate power demands and data processing and transmission requirements, which makes it particularly well suited for long-term monitoring and field campaigns. A promising direction for future research is coupling neutron observations to land surface models and possibly even inverting neutron data to obtain soil properties and evapotranspiration. Furthermore, advances in neutron detection technology, for example, in the area of directionally sensitive neutron detectors (Mascarenhas et al., 2006), have the potential to open new applications and spatial scales for hydrologic measurements.

Bibliography

- Chadwick, J., 1932. Possible existence of a neutron. *Nature*, **129**, 312.
- Desilets, D., Zreda, M., and Prabu, T., 2006. Extended scaling factors for in situ cosmogenic nuclides: new measurements at low latitude. *Earth and Planetary Science Letters*, **246**, 25.
- Famiglietti, J. S., Ryu, D., Berg, A. A., Rodell, M., and Jackson, T. J., 2008. Field observations of soil moisture variability across scales. *Water Resources Research*, **44**, W01423.
- Gardner, W., and Kirkham, D., 1952. Determination of soil moisture by neutron scattering. *Soil Science*, **73**, 391.
- Glasstone, S., and Edlund, M. C., 1952. *Elements of Nuclear Reactor Theory*. Van Nostrand: Princeton, pp. 148–150.
- Kodama, M., Kudo, S., and Kosuge, T., 1985. Application of a mospheric neutrons to soil moisture measurement. *Soil Science*, **140**, 237.
- Krane, K. S., 1987. *Introductory Nuclear Physics*. New York: Wiley.
- Kuwabara, T., Bieber, J. W., Clem, J., Evenson, P., et al., 2006. Real-time cosmic ray monitoring system for space weather. *Space Weather: The International Journal of Research and Applications*, **4**, S08001.
- Mascarenhas, N., Brennan, J., Krenz, K., Lund, J., et al., 2006. Development of a neutron scatter camera for fission neutrons.

In *IEEE Nuclear Science Symposium*, Vol. 1, San Diego, CA, pp. 185–188.

- Pelowitz, D. B. (ed.). 2005. *MCNPX User's Manual, Version 5*. LA-CP-05-0369, Los Alamos National Laboratory.
- Simpson, J. A., 1951. Neutrons produced in the atmosphere by the cosmic radiations. *Physics Reviews*, **83**, 1175.
- Uchiyama, Y., Aharonian, F. A., Tanaka, T., Takahashi, T., and Maedo, Y., 2007. Extremely fast acceleration of cosmic rays in a supernova remnant. *Nature*, **449**, 576.
- Western, A. W., and Blöschl, G., 1999. On the spatial scaling of soil moisture. *Journal of Hydrology*, **217**, 203.
- Zreda, M., Desilets, D., Ferré, T. P. A., and Scott, R. L., 2008. Measuring water content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophysical Research Letters*, **35**, L21402.

COST BENEFIT ASSESSMENT

Molly Macauley

Resources for the Future, Washington, DC, USA

Synonyms

Cost savings; Economic benefit; Net benefit; Societal benefit; Value of information

Definition

The “costs” of remote sensing usually refer to the direct monetary costs of designing, testing, constructing, deploying, operating, and maintaining the hardware (instruments, platforms, supporting infrastructure, communications networks) of remote sensing devices. These are typically the cameras, radar, lidar, or other instruments carried on aircraft or spacecraft for purposes of observing Earth processes (in situ remote sensing involves the hand-carrying of these devices). Costs also include the expenses of designing, testing, operating, and maintaining the software and other tools associated with using the data acquired from remote sensing instruments. Additional costs include those of personnel and may range from salaries of administrative and engineering experts to training and salaries of scientific researchers and the final “consumers” who use remote sensing for decision making. “Benefits” refer to the enhanced knowledge or scientific understanding gained from remote sensing as well as the practical applications which remote sensing data may contribute to. Benefits may be expressed in monetary measures or other quantitative measures such as “a percentage improvement in accuracy” or qualitative descriptions. “Cost benefit assessment” links costs and benefits in an attempt to measure the net gain from remote sensing – what benefits does it provide, after accounting for the costs incurred to glean those benefits?

Introduction

A decision whether to invest in remote sensing can be informed by understanding its costs and benefits. These costs and benefits are relevant to many decision makers, ranging from public officials who invest in remote sensing systems to consumers of information, such as managers of natural and environmental resources and the public at large.

Discussion

Costs

Although measuring the costs of remote sensing systems may be thought to be straightforward, accounting for all costs may be difficult in practice. The costs of hardware, software, and personnel may not be easily identified if they are included within broad categories in some cost accounting systems, making it difficult to assign costs specifically to a remote sensing system. For many users of remote sensing, data may be provided freely or at a marginal cost of reproduction, in which case costs as reported in budget statements may not represent the full cost to society of investment in a remote sensing system. In addition, the costs incurred to use remote sensing data may include common resources, such as shared computers, communications devices, and personnel, for which the allocation of cost among uses of shared facilities is difficult. In a study of how remote sensing data are transformed into information for managers of natural and environmental resources, such as national or local agencies or agricultural or geologic exploration companies, the National Research Council (NRC, 2001) identifies a range of indirect costs of transforming “data” into useful “information.” One of the conclusions of the NRC (p. 16) is that “transforming technical data into a form that is meaningful to nontechnical users – a process often including either the integration of remote sensing data with other types of data or scientific research to characterize the data (or both) – is highly dependent on the information requirements of applied users and on the skills of technical experts.” The NRC concludes that this process is poorly developed and an impediment – a cost – in effective use of remote sensing.

Benefits

Benefits of remote sensing include enhanced scientific understanding of the Earth and Earth processes, improved ability to make decisions in managing natural and environmental resources, and, arguably, the prestige accorded demonstration of technological prowess in building and deploying a remote sensing system or even the benefits to national or regional security associated with remote sensing of Earth resources. Benefits can accrue to the private sector as well as the public sector; in fact, some remote sensing systems are owned and operated exclusively by the private sector or as

joint arrangements between the private and public sectors.

The National Research Council (NRC, 2008) discusses the enhanced scientific understanding of the Earth and Earth processes from remote sensing. These accomplishments (benefits) include monitoring global [stratospheric ozone](#) depletion, detecting tropospheric ozone, measuring the Earth’s radiation budget, generating synoptic weather imagery, assimilating data for sophisticated numerical weather prediction, discovering the dynamics of ice sheet flows, detecting mesoscale variability of [ocean surface topography](#) and its importance in ocean mixing, observing the role of the ocean in climate variability, monitoring agricultural lands for a famine early warning system, and determining the Earth reference frame with unprecedented accuracy (see NRC, 2008, Table S.1, p. 4).

Numerous case studies have examined the benefits of applications of remote sensing for managing environmental and natural resources (for instance, see Battese, 1988; Ning, 1996; Nelson, 1997; Dudhani, 2006; Ward, 2000). Most of these studies express the benefits of remote sensing in terms of percentage improvements in statistical measures of accuracy; fewer studies express benefits in monetary terms. The primary difficulty in attempting to monetize benefits of remote sensing is that the environmental and natural resources themselves are typically “public goods,” that is, resources which, unlike ordinary goods and services, are not exchanged in markets (see Macauley, 2006 for a discussion of the challenges in valuing “information” such as remote sensing). Accordingly, it is difficult to ascribe monetary value to remote sensing data about nonmarketed resources. Bouma and coauthors (2009) carry out a case study of the value of satellite remote sensing of water quality in the North Sea and are able to monetize benefits in terms of the costs saved by water managers. Williamson and coauthors describe the economic value of remote sensing in improving forecasts of natural disasters in terms of reducing loss of life, property damage, and other costs. These approaches thus illustrate ways in which to express benefits in financial terms, but this line of research remains small and fragmented. And, because remote sensing data can often serve multiple purposes (remote sensing of land use, for instance, can inform highway planning as well as forecasts of agricultural production), a “unit” of data may be undervalued if only one of many applications of the data is evaluated in a case study.

Assessing costs and benefits

Few studies have attempted to assess the net benefits – that is, comparing both the costs and benefits of remote sensing and quantifying the extent to which benefits exceed costs. Given the difficulty in monetizing benefits, quantitatively relating benefits and costs is obviously even more difficult. In the few instances when benefits and costs are expressed in common units (say, dollars), the

comparison is easier. For example, if benefits are expressed in terms of costs saved, and if cost data are available, then benefits and costs can be compared (Bouma and coauthors make this comparison, for instance). If benefits and costs are combined, a general rule of thumb in benefit and cost assessment is that the *difference* between benefits and costs is preferred to the ratio between benefits and costs. Using the difference avoids the problem of whether a benefit is a negative cost, which can lead to ambiguous results when using the ratio rather than the numerical difference. (An example: In estimating the benefits of remote sensing in monitoring air pollution, is a reduction in pollution a benefit or an avoided cost? Expressing the reduction as a benefit or a cost will not affect the difference but will affect the result if benefits and costs are expressed as a ratio.)

Conclusion

Assessing the benefits and costs of remote sensing is one of the challenges of ascertaining the appropriate amount of investment a society should undertake in remote sensing systems. When is the cost of these systems justified by the benefits they confer? For a variety of reasons, quantifying costs and benefits is difficult. Expressing benefits in financial terms is particularly difficult, as remote sensing data may confer benefits in the form of new knowledge or about natural resources (air, water, climate, land, oceans) and the environment (air and water quality, land use). Society values these benefits, but ascribing monetary value to them is quite difficult.

Bibliography

- Battese, G. E., 1988. An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**(401), 28–36.
- Bouma, J. A., van der Woerd, H. J., and Kuik, O. J., 2009. Assessing the value of information for water quality management in the North Sea. *Journal of Environmental Management*, **90**(2), 1280–1288.
- Dudhani, S., 2006. Assessment of small hydropower potential using remote sensing data for sustainable development in India. *Energy Policy*, **34**(17), 3195–3205.
- Macauley, M. K., 2006. The value of information: measuring the contribution of space-derived earth science data to resource management. *Space Policy*, **22**, 274–282.
- National Research Council, Committee on Scientific Accomplishments of Earth Observations from Space, 2008. *Earth Observations from Space: The First 50 Years of Scientific Accomplishments*. Washington, DC: The National Academies.
- National Research Council, Space Studies Board and Ocean Studies Board, 2001. *Transforming Remote Sensing Data into Information and Applications*. Washington, DC: The National Academies.
- Nelson, G. C., 1997. Do roads cause deforestation? Using satellite images in econometric analysis of land use. *American Journal of Agricultural Economics*, **79**(1), 80–88.
- Ning, S. K., 1996. Soil erosion and non-point source pollution impacts assessment with the aid of multi-temporal remote sensing images. *Journal of Environmental Management*, **79**(1), 88–101.
- Ward, D., 2000. Monitoring growth in rapidly urbanizing areas using remotely sensed data. *The Professional Geographer*, **52**(3), 371–386.
- Williamson, R. A., Hertzfel, H. R., Cordes, J., and Logsdon, J. M., 2002. The socioeconomic benefits of earth science and applications research: reducing the risks and costs of natural disasters in the USA. *Space Policy*, **18**(1), 57–66.

Cross-references

[Commercial Remote Sensing](#)
[Data Policies](#)
[Emerging Applications](#)
[Environmental Treaties](#)
[Mission Operations, Science Applications/Requirements](#)

CROP STRESS

Susan Moran

USDA ARS Southwest Watershed Research Center,
 Tucson, AZ, USA

Synonyms

Insect infestation; Nitrogen deficiency; Water deficiency;
 Weed infestation

Definitions

Crop stress. Crop response to environmental factors that results in suboptimal crop production.

Introduction

Crop stress is the plant response to environmental factors that ultimately results in suboptimal crop production. The environmental factors of primary interest to US corn, cotton, soybean, and wheat producers are water, nutrients, weeds, and insects. Not coincidentally, these are also the factors that are most easily managed through irrigation and applications of fertilizer, herbicides, and pesticides. Crops are generally managed to minimize crop stress within the constraints of producing a profitable yield and minimizing environmental impact. The day-to-day management decisions to achieve this delicate balance are based in part on information about the extent, duration, and cause of crop stress. The role of remote sensing in crop management is to provide such information about crop stress using sensors that acquire data in the visible (VIS), near-infrared (NIR), short-wave infrared (SWIR), thermal infrared (TIR), and synthetic aperture radar (SAR) wavelengths. A first step is to understand the physical plant manifestations associated with crop stress that are most easily detected with optical and microwave remote sensing.

Plant manifestations of crop stress

Water stress affects the plant leaf canopy by two primary mechanisms (Rosenthal et al., 1987). The first involves the closure of leaf stomata, which results in a reduction in

photosynthesis and transpiration. The associated increase in leaf temperature can be detected using remote sensing in the thermal infrared wavelengths (Jackson et al., 1981). The second mechanism involves a decrease in leaf expansion and an increase in leaf senescence. The reduction in plant leaf canopy development in comparison to well-watered plants can be detected through estimates of LAI or ground cover made using remote sensing in the visible wavelengths (Maas and Rajan, 2008). The mechanism affecting leaf stomata is initiated when available soil water in the root zone falls below 30 %. In contrast, the mechanism affecting leaf expansion and senescence is initiated when available soil water in the root zone falls below 50 %. Thus, leaf expansion and senescence typically are affected by water stress before photosynthesis and transpiration. There are a number of secondary effects of water stress that are associated with the adaptation of plants to the decrease in water availability.

The crop physiological adaptations to transient water deficit range from changes in canopy architecture to adjustments in leaf osmotic potential (Turner, 1977). Many of these adaptations have a pronounced effect on spectral reflectance and SAR backscatter and the optical properties of plants that allow stress detection with remote sensing. Crops have the capacity for developmental plasticity to complete the life cycle before serious water deficits develop. For example, studies have shown that wheat can hasten maturity in response to mild water deficits at the critical time between flowering and maturity. To endure prolonged water deficit while maintaining high water potential, some crops reduce water loss through increased epidermal waxes of leaves and a reduction in general plant productivity. Other adaptations are to reduce the radiation absorbed by the plant through leaf movement (e.g., leaf cupping, paraheliotropism, or wilting) or to reduce leaf area through decreased leaf expansion, reduced tillering and branching, and leaf shedding. It is generally reported that leaves under water stress show a decrease in reflectance in the NIR spectrum and a reduced red absorption in the chlorophyll active band (0.68 μm); however, Guyot et al. (1984) found that it was necessary to have an extremely severe water stress to affect the leaf reflective properties. In the TIR, there is a direct link between the process of plant water evaporation and the plant thermal response (i.e., water evaporates and cools the leaves) explained by Jackson et al. (1981).

Like crop water stress, crop nutrient stress has a direct effect on crop growth and development. Nitrogen is frequently the major limiting nutrient in agricultural soils. Leaves deficient in nitrogen absorb less and scatter more visible light (Schepers et al., 1996). Due to the link between leaf chlorophyll and nitrogen concentrations (Daughtry et al., 2000), leaves marginally deficient in nitrogen may appear a lighter, less saturated shade of green, and more severely nitrogen-stressed leaves may appear yellowish green and chlorotic. Thomas and Oerther (1972) found that with nitrogen deficiency, the visible reflectance increased (due to decreasing

chlorophyll content) and the NIR and SWIR reflectances decreased (due to decreasing number of cell layers). Nutrient deficiencies in crop canopies have the potential to affect canopy architecture and the optical properties of not only the leaf but also the stem and flower/grain head. Manifestations of typical nutrient stresses generally appear initially as changes in the optical properties of leaves and only later as change in the canopy architecture and decreased canopy biomass. The position of the red edge (an abrupt, step increase in the leaf reflectance in the NIR around 0.72 μm just outside the visible region) offers a robust metric for monitoring leaf and canopy nutrient status (Peñuelas and Filella, 1998). Recently, the concentration of epidermal polyphenolics, secondary metabolites in the leaf that may be measured using a commercially available clip-on UV absorption meter, has shown promise as a surrogate measure of leaf nitrogen status (see, for example, Tremblay et al., 2007; Demotes-Mainard et al., 2008; Meyer et al., 2006).

Crop stress due to weed interference has been attributed to many factors, including allelopathy and competition for sunlight, soil water, and nutrients (Sikkema and Dekker, 1987). The plant manifestation of weed-induced crop stress is generally reduced crop yields. Because weed distribution is influenced by drainage, topography, soil type, and microclimate, crop stress in weed-infested fields is highly variable. Variations in reflectance patterns and canopy temperatures over time and space may reveal crop stress associated with soil and topographic conditions (Wiles et al., 1992). In the early season, herbicide application is based simply on the presence or absence of plants, and remote sensing systems generally use the reflectance differences between relatively wide spectral bands in the visible and NIR spectra to make the distinction between plants and soil or rock (Medlin et al., 2000). Post-emergent herbicide applications require discrimination between weeds and crops, which is generally accomplished by using the difference between spectral signatures of crops and specific weeds or by acquiring images when weed coloring is particularly distinctive (Brown et al., 1994) or weed patches are comparatively large, dense, and/or tall (Pérez et al., 2000).

Remote sensing is not used to directly observe insects, but rather, to observe the damage to crop foliage and to detect plant canopy conditions that might be conducive to insect infestation. Early infestations of some insects are associated with leaf senescence or mortality, resulting in reduction in canopy density. Some crop pests not only cause physical damage to the leaf canopy but also cause a change in the spectral reflectance characteristics of the affected foliage. Aphids (*Aphididae*) deposit honeydew on cotton leaves which supports the growth of sooty mold (*Aspergillus* spp.), thus profoundly affected the reflectance characteristics of the leaves, particularly in the NIR (Maas, 1998). Other pests, such as spider mites (*Tetranychus* spp.), can cause changes in leaf reflectance that can be detected using remote sensing in the visible and NIR wavelengths (Fitzgerald et al., 2001, 2004). These spectral

signatures are distinct enough to differentiate them from water stress effects (Fitzgerald et al., 2000). Insects are sometimes attracted to areas within fields that contain the most vigorous plant growth. Areas of lush cotton growth in Louisiana were identified with spectral vegetation indices to direct scouting for the tarnished plant bug (*Lygus lineolaris*) and facilitate spatially variable insecticide applications (Willers et al., 2000).

Remote sensing of crop stress

Readers are referred to recent reviews by Moran et al. (2004) and Hatfield et al. (2004) for an extensive summary of remote sensing applications and products for detecting crop stress associated with water and nutrient deficiencies and weed and insect infestations. The greatest progress has been made in crop water stress detection. The use of remote sensing in irrigation scheduling has been reviewed by Maas (2003). Some spectral crop water stress indices have been commercialized for irrigation scheduling (Jackson et al., 1981; Burke et al., 1988). Opportunities for deriving crop nutrient status and pest infestation from remote sensing have recently increased with the development of hyperspectral and narrowband multispectral imaging sensors (Gitelson et al., 2006; LaCapra et al., 1996). The production of fine-resolution digital elevation models (DEM) with high vertical accuracies from radar systems provides useful supplemental information for the management of large agricultural areas. Recent development and launches of multispectral sensors with fine resolution has stimulated efforts to observe the early stages of pest infestations and areas with potential for pest infestations in time for control measures (Fitzgerald et al., 2004).

Conclusions

The technologies of the future will probably include sensors to measure natural and genetically induced fluorescence related to crop vigor (e.g., Liu et al., 1997), more focus on multispectral data fusion including SWIR, TIR, and microwave measurements (e.g., Jackson et al., 2004), and increased assimilation of remotely sensed data in crop yield models and decision support systems (e.g., Maas, 2005; Baez et al., 2005; Maas, 2005; Ko et al., 2005, 2006). The latter has the potential to address one of the greatest challenges to use of remote sensing as a source of information about crop stress – determining the cause of crop stress. It has been particularly difficult to discriminate crop stress due to water and nitrogen, which often produce similar plant manifestations but require different and costly management. Decision support systems based on crop growth theory can assimilate producer knowledge, management history, and remote sensing information to best determine the extent, magnitude, and cause of crop stress. This could lead to the turnkey solution called for by Hatfield et al. (2008) for application of remote sensing for agronomic decisions suited to both specialists and nonspecialists.

Bibliography

- Baez, A. D., Kiniry, J., Maas, S. J., Tiscereno, M., Macias, J., Mendoza, J., Richardson, C., Salinas, G. J., and Manjarrez, J. R., 2005. Large-area maize yield forecasting using leaf area index based yield model. *Agronomy Journal*, **97**(2), 418–425.
- Brown, R. B., Steckler, J.-P. G. A., and Anderson, G. W., 1994. Remote sensing for identification of weeds in no-till corn. *Transactions of ASAE*, **37**, 297–302.
- Burke, J. J., Mahan, J. R., and Hatfield, J. L., 1988. Crop-specific thermal kinetic windows in relation to wheat and cotton biomass production. *Agronomy Journal*, **80**, 553–556.
- Daughtry, C. S. T., Walthall, C. L., Kim, M. S., deColstoun, E. B., and McMurtrey, J. E., 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, **74**, 229–239.
- Demotes-Mainard, S., Boumaza, R., Meyer, S., and Cerovic, Z. G., 2008. Indicators of nitrogen status for ornamental woody plants based on optical measurements of leaf epidermal polyphenol and chlorophyll contents. *Scientia Horticulturae*, **115**, 377–385.
- Fitzgerald, G. J., Maas, S. J., and DeTar, W. R., 2000. Assessing spider mite damage in cotton using multispectral remote sensing. In *Proceedings of the Beltwide Cotton Conference*, San Antonio, TX, Vol. 2, pp. 1342–1345.
- Fitzgerald, G. J., Maas, S. J., and DeTar, W. R., 2001. Spying on spider mites from on high. *California-Arizona-Texas Cotton Magazine*, **Fall**, 2001, 13–15.
- Fitzgerald, G. J., Maas, S. J., Pinter, P. J., Hunsaker, D. J., and Clarke, T. R., 2002. Spider mite damage mapping and cotton canopy biophysical characterization using spectral mixture analysis. In *Proceedings Symposium on American-Taiwan Agricultural Cooperative Projects*, Taipei, Taiwan, November 15, 2005.
- Fitzgerald, G. J., Maas, S. J., and DeTar, W. R., 2004. Spider mite detection and canopy component mapping in cotton using hyperspectral imagery and spectral mixture analysis. *Precision Agriculture*, **5**, 279–289.
- Gitelson, A. A., Keydan, G. P., and Merzlyak, M. N., 2006. Three-band model for noninvasive estimation of chlorophyll, carotenoids and anthocyanin contents in higher plant leaves. *Geophysical Research Letters*, **33**, 111402, doi:10.1029/2006G1026457.
- Guyot, G., Dupont, O., Joannes, H., Prieur, C., Huet, M., and Podaire, A., 1984. Investigation into the Mid-IR spectral band best suited to monitoring vegetation water content. In *Proceedings of 18th International Symposium on Remote Sensing of the Environment*, October 1–5, 1984, Paris, pp. 1049–1063.
- Hatfield, J. L., Prueger, J. H., Kustas, W., Hatfield, J. L., Prueger, J. H., and Kustas, W. P., 2004. In Ustin, S. (ed.), *Remote Sensing for Natural Resources Management and Environmental Monitoring: Manual of Remote Sensing*, 3rd edn. New York: Wiley, Vol. 4, pp. 531–568. Chap. 10.
- Hatfield, J. L., Gitelson, A. A., Schepers, J. S., Schepers, J. S., and Walthall, C. L., 2008. Application of spectral remote sensing for agronomic decisions. *Agronomy Journal*, **100**, S-117–S-131.
- Jackson, R. D., Idso, S. B., Reginato, R. J., and Pinter, P. J., Jr., 1981. Canopy temperature as a crop water stress indicator. *Water Resources Research*, **17**, 1133–1138.
- Jackson, T. J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., Doraiswamy, P., and Hunt, E. R., 2004. Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. *Remote Sensing of Environment*, **92**, 475–482.
- Ko, J., Maas, S. J., Lascano, R. J., and Wanjura, D., 2005. Modification of the GRAMI model for cotton. *Agronomy Journal*, **97**(5), 1374–1379.
- Ko, J., Maas, S. J., Mauget, S., Piccinni, G., and Wanjura, D., 2006. Modeling water-stressed cotton growth using within-season remote sensing data. *Agronomy Journal*, **98**, 1600–1609.

- LaCapra, V. C., Melack, J. M., Gastil, M., and Valeriano, D., 1996. Remote sensing of foliar chemistry of inundated rice with imaging spectrometry. *Remote Sensing of Environment*, **55**, 50–58.
- Liu, C., Muchhal, U. S., and Raghothama, K. G., 1997. Differential expression of TPS-11, a phosphate starvation-induced gene in tomato. *Plant Molecular Biology*, **33**, 867–874.
- Maas, S. J., 1998. Remote sensing: value in a bird's-eye view. In Wrona, A. F. (ed.), *Cotton Physiology Today*, 9th edn. Memphis: National Cotton Council, Vol. 2, pp. 13–15.
- Maas, S. J., 2003. Irrigation scheduling by remote sensing technologies. In Stewart, B. A., and Howell, T. A. (eds.), *Encyclopedia of Water Science*. New York: Marcel Dekker, pp. 523–527.
- Maas, S. J., 2005. An overview of interfacing models with remote sensing. In *Proceedings of the 35th Biological Systems Simulation Conference*. Phoenix, AZ, pp. 31–33.
- Maas, S. J., and Rajan, N., 2008. Estimating ground cover of field crops using medium-resolution multispectral satellite imagery. *Agronomy Journal*, **100**(2), 320–327.
- Maas, S., Lascano, R., Cooke, D., Richardson, C., Upchurch, D., Wanjura, D., Krieg, D., Mengel, S., Ko, J., Paynem, W., Rush, C., Brightbill, J., Bronson, K., Guo, W., and Rajapakse, S., 2004. Within-season estimation of evapotranspiration and soil moisture in the high plains using YieldTracker. In *Proceedings of the 2004 High Plains Groundwater Resources Conference*. Lubbock, TX, pp. 219–226.
- Medlin, C. R., Shaw, D. R., Gerard, P. D., and LaMastus, F. E., 2000. Using remote sensing to detect weed infestations in Glycine max. *Weed Science*, **48**, 393–398.
- Meyer, S., Cerovic, Z. G., Goulas, Y., Montpied, P., Demotes-Mainard, S., Bidel, L. P. R., Moya, I., and Dreyer, E., 2006. Relationships between optically assessed polyphenols and chlorophyll contents, and leaf mass per area ratio in woody plants: a signature of the carbon-nitrogen balance within leaves? *Plant, Cell & Environment*, **29**, 1338–1348.
- Moran, M. S., Maas, S. J., Vanderbilt, V. C., Barnes, E. M., Miller, S. N., and Clarke, T. R., 2004. Application of image-based remote sensing to irrigated agriculture. In Ustin, S. (ed.), *Remote Sensing for Natural Resources Management and Environmental Monitoring: Manual of Remote Sensing*, 3rd edn. New York: Wiley, Vol. 4, pp. 617–676. Chap. 12.
- Peñuelas, J., and Filella, I., 1998. Visible and near infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science*, **3**(4), 151–156.
- Pérez, A. J., López, F., Benlloch, J. V., and Christensen, S., 2000. Colour and shape analysis techniques for weed detection in cereal fields. *Computers and Electronics in Agriculture*, **25**, 197–212.
- Rosenthal, W., Arkin, G., Shouse, P., and Jordan, W., 1987. Water deficit effects on transpiration and leaf growth. *Agronomy Journal*, **79**, 1019–1026.
- Schepers, J. S., Blackmer, T. M., Wilhelm, W. W., and Resende, M., 1996. Transmittance and reflectance measurements of corn leaves from plants with different nitrogen and water supply. *Journal of Plant Physiology*, **148**, 523–529.
- Sikkema, P. H., and Dekker, J., 1987. Use of infrared thermometry in determining critical stress periods induced by quackgrass (*Agropyron repens*) in Soybeans (*Glycine max*). *Weed Science*, **35**, 784–791.
- Thomas, J. R., and Oerther, G. F., 1972. Estimating nitrogen content of sweet pepper leaves by reflectance measurements. *Agronomy Journal*, **64**, 11–13.
- Tremblay, N., Wang, Z., and Belec, C., 2007. Evaluation of the DualEx for the assessment of corn nitrogen status. *Journal of Plant Nutrition*, **30**, 1355–1369.
- Turner, N. C., 1977. Drought resistance and adaptation to water deficits in crop plants. In Mussell, H., and Staples, R. C. (eds.), *Stress Physiology in Crop Plants*. New York: Wiley, pp. 344–372.
- Wiles, L. J., Oliver, G. W., York, A. C., Gold, H. J., and Wilkerson, G. G., 1992. Spatial distribution of broadleaf weeds in North Carolina soybean (*Glycine max*) fields. *Weed Science*, **40**, 554–557.
- Willers, J. L., Seal, M. R., and Luttrell, R. G., 1999. Remote sensing, line-intercept sampling for tarnished plant bugs (Heteroptera: Miridae) in mid-South cotton. *Journal of Cotton Science*, **3**, 160–170.

Cross-references

[Agriculture and Remote Sensing](#)
[Data Assimilation](#)
[Irrigation Management](#)
[Precision Agriculture](#)
[Soil Moisture](#)
[Soil Properties](#)
[Vegetation Indices](#)
[Vegetation Phenology](#)

CRYOSPHERE AND POLAR REGION OBSERVING SYSTEM

Mark Drinkwater

Mission Science Division, European Space Agency,
 ESA/ESTEC, Noordwijk ZH, The Netherlands

Definition

Cryosphere. It collectively describes elements of the Earth system containing water in its frozen state and includes sea ice, lake and river ice, snow cover, solid precipitation, glaciers, ice caps, ice sheets, permafrost, and seasonally frozen ground. Although a significant portion of the world's snow and ice is found in the polar regions, cryosphere exists at all latitudes and in about 100 countries.

Polar regions. Earth's polar regions are the areas of the globe surrounding the north and south poles typically encompassed by a line of latitude corresponding to 66° 34' north or south (i.e., between each pole and its corresponding polar circle), which is the approximate limit of the midnight sun and the polar night. Alternatively, it can be defined as the region where the average temperature for the warmest month (July) is below 10 °C (50 °F).

Introduction

Knowledge of the state of the cryosphere is important for weather and climate prediction, assessment and prediction of [sea level rise](#), availability of freshwater resources, navigation, shipping, fishing, mineral resource exploration and exploitation, and in many other practical applications (IGOS, 2007). Changes to the cryosphere have far-reaching climate and socioeconomic consequences. The need for reliable global monitoring is essential to address the issues of climate and cryosphere within the Earth system. Despite its importance, the cryosphere remains one of

the most under-sampled domains in the Earth's climate system.

The *Cryosphere and Polar Region Observing System* (hereafter abbreviated as *CryOS*) shall provide a stable, long-term monitoring capability using a combination of in situ and remote sensing measurement capabilities (Drinkwater et al., 2008). The satellite remote sensing part of this system specifically fulfills the need for regular measurements of changes in the global cryosphere in response to climate variability and change. In particular, due to the logistical challenge of accessing the remote polar regions and high elevation cryosphere, together with the harsh weather conditions and extended winter-season darkness, satellites are the primary means of making year-round, day and night, weather-independent observations. Their advantage also lies in their ability to obtain uniform, consistent global sampling, including observations of the north and south polar ice-covered regions and lower-latitude mountain glaciers, permafrost, lake and river ice, and seasonally snow-covered areas.

The objectives of the satellite element of *CryOS* shall be to quantify snow- and ice-mass variability and mass, energy, freshwater, and gas exchanges between the cryosphere and the other elements of the Earth system (land, atmosphere, and ocean). This system element must respond to climate research and policy needs, socioeconomic needs, and operational snow and ice service needs. Moreover, the resulting data shall be made easily and freely available and shall be integrated with data from airborne and in situ observing systems for the purpose of developing a comprehensive record on cryospheric variability (Drinkwater et al., 2008, 2010).

Satellites contributing to *CryOS* must comprise a broad range of capabilities and have the ability both to sample the primary aspects of the cryosphere. Due to the sensitivity of the cryosphere to temperature and corresponding short timescales on which snow and ice may change, the system must be capable of providing global coverage with daily sampling from broad-swath, low-resolution sensors, complemented by high-resolution satellite sensors with the ability to provide detailed local and regional information, as required by the specific application.

Essential satellite elements of an observing system

The cryosphere and polar observing system must contain the following basic satellite capabilities:

- Daily, all-weather, global imaging
- High-resolution all-weather polar ice dynamics imaging
- Surface temperature and albedo change
- Ice topography/elevation/thickness change
- Gravity and mass distribution and exchange
- Position monitoring/navigation
- Telecommunications for data relay

Additionally, the geophysical products described in the list above require a source of independent measurement

data for product validation. Together these general capabilities are required to provide robust information on the main terrestrial and marine cryospheric domains.

Terrestrial cryosphere

For the terrestrial cryosphere, *CryOS* shall provide a complete picture of the snow reserves and solid precipitation, river and lake ice, permafrost, seasonally frozen ground, glaciers, ice caps, and ice sheets. This element of the observing system bridges meteorological and hydrological applications and ensures incorporation of data on the appropriate cryospheric variables in the next generation of hydrological and climate models. The basic challenge is to provide global data with the accuracy required for water management and disaster mitigation (e.g., avalanches), for monitoring climate change and variability, and for the estimation of [sea level rise](#).

Marine cryosphere

For the marine cryosphere, *CryOS* shall provide regular observations of sea ice and ice shelf characteristics and their dynamics. Ideally these data shall be supported by Global Ocean Observing System (GOOS) observations of ocean temperature, salinity, dynamic topography, and tides to understand the ocean-ice-atmosphere coupling processes responsible for variability and changes.

The following sections treat each of the cryospheric subdomains which require specific monitoring capabilities.

Snow extent and water equivalent

Terrestrial snow cover has the largest geographic extent of the components of the cryosphere covering nearly 50 million km² of the Northern Hemisphere in winter. Snow, with its high albedo, influences surface water and energy fluxes, atmospheric dynamics and weather, frozen ground and permafrost, biogeochemical fluxes, and ecosystem dynamics.

In order to support water, weather, and climate applications, various observations of snow are required. Remote sensing observations focus on the daily geographic extent of snow cover and snow water equivalent (SWE; the total water content) since these two quantities have a fundamental control on hydrologic and ecosystem processes.

Point measurements of solid precipitation and snow depth on the ground are often unrepresentative of surrounding areas, and so satellite remote sensing data must be combined with these snow measurement sites to provide a consistent picture at regional or continental scale. Satellite measurements of snow extent are typically accomplished using visible and near-infrared sensors on satellites in either polar low Earth orbit (LEO) or geostationary orbit (GEO). Using such multispectral measurements from GOES, MSG/SEVIRI, AVHRR, MODIS, MERIS, MISR, and AATSR, it is possible to discriminate between snow- and cloud-covered regions and to establish

the extent of snow cover. Currently, these sensor data are complemented by all-weather SWE products produced from passive and active microwave data with appropriate frequencies such as SSM/I, AMSR-E, and ASCAT; QuikSCAT; and Oceansat-2 Scat though their low spatial resolution currently limits their use in hydrometeorological models. Future multiple-frequency (Ku- and X-band), high-resolution active microwave sensors such as synthetic aperture radars (SAR) (see entry [Data Processing, SAR Sensors](#)) must be developed specifically for measurement of SWE at the relevant sub-kilometer scale.

Solid precipitation

To date satellite precipitation measurements have greatly increased our ability to monitor and observe liquid precipitation (i.e., rainfall) globally. However, a similar capability does not yet exist at high latitudes for solid precipitation (i.e., snow and hail). Thus, development of a robust snowfall measurement capability remains a high priority for the cryosphere, such that the contribution of this critical element of the high-latitude water cycle can be fully characterized.

River and lake ice

Lake and river ice are a key component of the terrestrial cryosphere, and seasonal ice growth plays an important role in regulating physical, chemical, and biological processes. Long lake and river ice records serve as an indicator of high-latitude climate variability and serve as an important data source for initializing weather forecast and climate models (Jeffries et al., [in press](#)). Meanwhile, the presence of freshwater ice also has a number of socioeconomic implications ranging from transportation to the occurrence and damaging effects of ice-jam flooding.

The observing system must be able to monitor the seasonal distribution and duration of lake and river ice. Autumn freeze-up and spring time thaw may be monitored at high resolution (10–120 m) under daylight conditions using Landsat MSS or TM and SPOT HRV for small lakes and narrow rivers, whereas for larger features the detection of the presence of ice has more traditionally been accomplished using medium-resolution (250 m–1 km) visible and infrared AVHRR and MODIS products. Nonetheless, extensive cloud cover and periods of darkness in winter limit their use at high latitudes. For establishing freeze-up and break-up dates and for year-round monitoring of seasonal ice on smaller lakes and rivers, SAR image products from ERS-1 and ERS-2, RADARSAT-1 and RADARSAT-2, Envisat ASAR, ALOS PALSAR, and TerraSAR-X may be considered more optimal (Jeffries et al., [2005](#)). These are complemented by scatterometer and passive microwave monitoring of the largest lakes.

Permafrost

Permafrost is subsurface soil which remains at or below the freezing point of water for 2 or more years.

Such permanently frozen ground and seasonally or intermittently frozen ground is widespread in the Arctic, subarctic, alpine, and high plateau regions, and in ice-free areas of the Antarctic and subantarctic. In the Northern Hemisphere, permafrost is estimated to cover approximately 23 million km² with up to 17 million km² of this underlying exposed land (Brown et al., [1997](#)). Seasonal and intermittently frozen ground together occupies approximately 56 million km² in the Northern Hemisphere and includes the “active layer” over permafrost (i.e., the layer which undergoes seasonal thawing) and soils outside permafrost regions (Zhang et al., [2003](#)).

Permafrost and the freeze-thaw state of land surfaces exert a critical influence on the surface energy balance, hydrologic cycle, ecosystems, biogeochemical fluxes, hydrology, and weather and climate systems. Further, permafrost thaw can have significant socioeconomic consequences through its direct impact on structures such as roads, buildings, railways, and pipelines.

The primary needs in permafrost areas are to determine the terrain and its changes over time (including topography and vegetation characteristics), in association with the thermal state of the surface. Near surface soil freeze-thaw state is a critical variable since it is a primary indicator in relation to the fluxes of energy, water, and carbon. Unlike other components of the terrestrial cryosphere, subsurface properties of permafrost terrain are not directly observable from remote sensing platforms. However, many surface features of permafrost terrain are observable with a variety of sensors ranging from aerial photography and high-resolution optical satellite imagery to synthetic aperture radar and satellite passive microwave radiometry (Duguay et al., [2005](#); Kääb, [2008](#)).

Passive optical imaging radiometers operating at visible and infrared wavelengths have traditionally been used under cloud-free conditions to map changes in the characteristics and thermal condition of permafrost terrain. Very-high-resolution (0.2–5 m pixels), narrow swath optical sensors such as EO1 Hyperion, Ikonos, or QuickBird can provide extremely detailed information on small features of the landscape. For mapping the evolution of features over time, ASTER, ALOS PRISM, ALOS AVNIR-2, and SPOT image data (10–100 m pixels) can be used, with the advantage of availability of SPOT data since the mid-1980s. For large area land cover mapping, these are complemented by medium-resolution (250 m–1 km pixels) broader swath imaging instruments (such as MERIS, MODIS) providing regional to continental scale coverage on a daily basis. Low-resolution (12–25 km pixels) image data from passive microwave radiometers and scatterometers, for example, have the advantage of year-round, daily, weather-independent continental scale coverage, and may be used to monitor the circumpolar spatial and temporal details of freeze and thaw cycles.

In addition, digital terrain models (DTMs) are also essential for understanding and modeling permafrost and

its evolution. High-resolution and high vertical accuracy DTMs may be derived at high-latitude tundra permafrost from repeat-pass InSAR or tandem interferometry (i.e., ERS-2-Envisat ASAR cross interferometry, COSMO-SkyMed, or TanDEM-X) measurements. Meanwhile changes in permafrost terrain may be detected using differential InSAR or DInSAR (i.e., the calculation of ground deformation from the difference between two InSAR images). In addition to InSAR techniques, lower-precision optical DTMs can be derived from stereoscopic optical sensors, such as SPOT HRV, Cartosat-1, and ALOS PRISM, or using stereophotogrammetry.

Ice sheets (topographic change, mass change, and sea level)

Ice sheets play a key role in the climate system, locking up vast amounts of freshwater in the form of snow and solid ice, and by their significant influence on the energy balance of the polar regions. The two major ice sheets remaining from the last ice age blanket most of Greenland and Antarctica. They contain enough freshwater, in the form of ice, to raise sea level by approximately 7.2 and 62 m, respectively (Alley et al., 2005). The inland ice rests on bedrock, while their floating ice shelf extensions link the ice sheet with the ocean. Ice is transported from the inland toward the ice shelves via fast-flowing ice streams and outlet glaciers.

Ice sheets undergo seasonal growth by snowfall and then wasting by summer melting, with accompanying release of freshwater into the ocean. The extent to which ice sheets and glaciers are changing is reported as “mass balance” by measuring the net inputs and net losses. The totals of precipitation and accumulation, ablation, melting, runoff, sublimation/evaporation, ice dynamics, and iceberg calving are reported over the course of each year and used to compute the net balance either as a gain in mass (positive) or negative loss of mass (negative). Ice sheets may lose mass by a combination of dynamical processes which govern ice stream flow across the grounding line, or by surface ablation, basal melting, and iceberg calving at the seaward margins of floating ice shelves. Exchange of mass between ice sheets and oceans as freshwater has a major impact on both ocean circulation and sea level and is currently estimated to contribute more than 0.3 mm/year (30 cm/century) to sea level.

Since the early 1990s, there are indications that rapid changes are taking place particularly around the margins of the Greenland and Antarctic ice sheets (Rignot et al., 2008; Van Den Broeke et al., 2009). A combination of remote sensing methods have highlighted that streaming ice flow (i.e., ice motion) is responsible for the largest proportion of mass loss from the ice sheets. By contrast, the central plateau of both ice sheets appears to be close to net neutral balance, with accumulation balancing ice-mass losses.

Remote sensing of ice sheet behavior relies on the ability of the sensors to acquire data over vast areas of terrain,

during day or night and also in all-weather conditions. Optical satellite data at visible wavelengths have been used since the 1960s to define changes in the ice margin and surface characteristics, with pairs of images acquired at different times used for motion of the surface and digital terrain model generation, the latter in stereographic mode.

Passive microwave radiometer or scatterometer images of ice sheets are often used to measure the onset, duration, and extent of ice sheet surface melt, since they are sensitive to the appearance of melt water in snow. Each has also been used to make estimates of mean annual snow accumulation.

Changes in ice sheet surface elevation averaged over large areas can be interpreted as a measure of changing ice sheet volume. Since 1978, successive radar altimeters have measured ice sheet and ice shelf surface elevation. Their primary limitation has been their ability to recover accurate data in the steep terrain around the edges of the ice sheets. Recently, this limitation has been addressed with the CryoSat-2 SAR interferometric radar altimeter. The ICESat laser altimeter has also recently acquired detailed topography around the sloping margins of the Greenland and Antarctic ice sheets. The advantage of the laser altimeter is its smaller measurement spot on the surface, allowing more accurate elevation data, albeit limited by optically thick cloud conditions.

SAR measurements complement the other remote sensing instrument data described above, through high-resolution observations. Since the early 1990s, SAR images have provided a wealth of new information on characteristics of the ice sheet surface, though the advent of SAR interferometry (InSAR) has enabled measurement of details of topography as well as precise details on the motion of ice (Rignot et al., 2011). In conjunction with ice sheet models, InSAR data provide a valuable means to study the impact of ice stream velocity on the net loss of ice mass.

Satellite gravity provides direct measurements of gravity anomalies due to ice sheet mass unloading or ice sheet mass change. Data from the GRACE satellite confirms other measurements which suggest that the Greenland and Antarctic ice sheets are losing mass around their margins over the last decade (Rignot et al., 2008; Van den Broeke et al., 2009). The high-resolution geoid and gravity anomaly data acquired by the GOCE satellite will serve as a precise reference for elevation changes and to understand the details of gravity anomalies due to isostatic adjustments to ice-mass unloading.

The observing system for quantifying ice sheet and glacier changes and their contribution to sea level include a number of contributing measurement capabilities (Wilson et al., 2010). Continued surveys by satellite radar and laser altimeters are needed to provide elevation changes over broad areas on a routine basis. These should ideally be coupled with continued time series of interferometric synthetic aperture radar (InSAR) measurements of glacier velocities to understand changes in ice thickness due to ice discharge. In addition, time series of satellite

gravimetric measurements of the static and time-varying components of the gravity field reveal net ice-mass changes over large regions and entire ice sheets. Periodic surveys by optical systems may also be beneficial as another means of mapping of changes in ice stream margins, ice shelves, or for complementary Digital Elevation Model generation.

Sea ice extent and concentration

A significant fraction of the surface of the world's polar ocean is frozen at any given time. Sea ice grows seasonally from 7 to 16×10^6 km² in the Northern Hemisphere and from 3 to 19×10^6 km² in the Southern Hemisphere, covering up to about 10 % of the surface area of each hemisphere at the respective winter peak. Arctic sea ice observations over the last 30 years indicate significant decreases in total area and extent of sea ice, although regional patterns of change can be highly variable and indicate opposing patterns of growth or decay in extent on interannual to decadal timescales.

Polar sea ice has an important climate regulating impact by limiting exchanges of momentum, heat, and moisture between the ocean and atmosphere. Due to the large difference in albedo between ice and ocean, reductions in sea ice extent by melting result in more heat being absorbed by the ocean, thereby amplifying the effects of high-latitude warming. Sea ice redistributes salt and freshwater by rejecting brine when it freezes, transporting freshwater by ice drift, and ultimately by depositing this freshwater in the upper ocean during summer melt. Thus sea ice variability exerts a considerable impact on the regional energy and freshwater budgets, as well as directly impacting the temperature, salinity, and buoyancy of brine in the upper ocean.

Daily observations of sea ice extent and concentration (fractional coverage of ice in a given area) are required to understand the thermodynamic evolution of the sea ice cover on seasonal to interannual timescales (Breivik et al., 2010). Satellite passive microwave radiometer observations (from ESMR, SMMR, SSM/I, and AMSR-E) have traditionally been used for this purpose as they are capable of daily mapping of the entire sea ice cover of both polar regions. The 30 year continuous time series of dual-polarized, multifrequency passive microwave radiometer data spanning the period from 1979 until the present day (Cavalieri et al., 1996; Comiso et al., 2003) has become the backbone of the sea ice monitoring system (see <http://www.nsidc.org/seaice/>). More recently, radar scatterometer images have been used to complement these radiometer data (Long et al., 2001). Their combination helps to distinguish between seasonal and perennial, or multiyear, ice and to provide more robust information on ice conditions during surface melting conditions. This synergy allows more robust retrieval of sea ice characteristics in operational sea ice products (e.g., EUMETSAT sea ice products: <http://saf.met.no/p/ice>).

Daily passive microwave radiometer observations have been supplemented by continuous visible and infrared very-high-resolution radiometer (1 km pixels) data since the 1970s with the NOAA AVHRR instrument series and DMSP OLS. More recently, the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) added to this capability. During cloud-free, sunlit periods, this class of sensors provides supplemental information on ice concentration and ice characteristics such as albedo and thickness from reflective characteristics. During the cold season, cloud-free infrared data may also be used to derive ice surface temperature using a split-window technique which employs two thermal bands (typically at 11 and 12 μ m).

The advent of SAR has revolutionized the capability to study sea ice at much higher resolution. Since the early 1990s, SAR has provided detailed images of sea ice at spatial resolutions of up to 30 m, including information fracture or lead locations/orientations with relatively thin ice, and information on ice ridging with relatively thicker, deformed ice. Wide-swath SAR (up to 500 km wide swath) has added the capability to deliver synoptic-scale coverage (up to 1,000 km) on timescales of 1–3 days, with complete Arctic coverage on a weekly basis (Kwok et al., 1999).

Sea ice thickness and dynamics

In order to fully appreciate the variability in sea ice mass and volume in response to climate variability, it is also required to measure sea ice thickness and drift dynamics. Both variables are essential to partition the role of thermodynamic and dynamic contributions to ice thickness changes, or to quantify advective fluxes of freshwater in the form of ice.

Ice drift data are derived from image pairs of the same region typically spaced at intervals of one to several days, using a variety of computer tracking algorithms which measure displacement of ice features over time. Daily passive microwave radiometer and scatterometer products have been used as a basis for Arctic and Antarctic ice drift products (Breivik et al., 2010), with merging of the drift fields from these two instruments providing more reliable results over several day intervals (Laverne et al., 2010). Notably, the accuracy of these products is limited by the resolution of the sensor and the intervals between images, which prevents the details of ice drift to be recorded on short time (<2 days) and space (<10 km grid) scales. For this purpose wide-swath SAR data from RADARSAT and Envisat ASAR have been used to estimate ice drift on a much finer grid, limited only by the image resolution and pixel spacing of 100–150 m. Using consistent time series of such products, it is possible to resolve the details of ice drift velocity and divergence and convergence of the ice, together with the dynamical thickening and drift-related area flux of sea ice (Kwok, 2011).

Since data from active and passive microwave imaging sensors do not contain sea ice thickness information

content, it was necessary to develop this capability. Radar altimeter measurements of ice surface elevation or freeboard (i.e., the difference in elevation between the ice surface and sea level in leads), together with ice density and snow-loading information, allow conversion of freeboard to ice thickness along profiles across the ice surface. Both the ICESat lidar and CryoSat-2 radar altimeter missions, with their higher resolution and measurement precision, allow changes in thickness and volume of sea ice to be estimated over sea ice cover.

The final goal of combination of sea ice thickness, concentration, and dynamics data is to make a complete estimation of ice volume fluxes (Kwok et al., 2009; Kwok, 2011). Using time series of each of these quantities, it becomes possible to estimate the seasonal to interannual variability in ice fluxes and, thus, the variability in transport of freshwater from high to lower latitudes.

With the future combination of passive microwave radiometers such as SSM/I and JPSS/MIS, and C-band SAR such as RADARSAT-2 and RADARSAT-3, and GMES Sentinel-1, and SAR altimeters such as CryoSat-2 and GMES Sentinel-3, a robust sea ice observing system component is assured for the next decade or so.

Albedo and surface temperature measurements

Observations of broadband albedo and surface temperature are required for characterizing surface-atmosphere radiation budgets and energy exchanges over polar snow, ice, land, and ocean surfaces. Albedo varies considerably for different snow and ice surface properties (from 35 % to 95 %), and changes dramatically with snow metamorphism due to temperature variability or melting, or underlying surface properties (Perovich et al., 2002). Broadband albedo measurements are accomplished by passive, multi-angle optical measurements in the visible and shortwave-infrared spectral range.

Surface temperature of snow, ice, ocean, or frozen ground is also important for the determination of energy exchanges and for understanding onset of processes such as snow metamorphism or freeze and thaw. Accurate temperature observations help constrain surface radiative and turbulent fluxes, and can help improve water and energy budgets for the cryosphere.

To date these measurements have been accomplished by passive visible and infrared radiometers operating in the visible and infrared range such as AVHRR, ASTER, ATSR, AATSR, MODIS, MISR, and MERIS instruments, with the future continuity in measurements secured with the future GMES Sentinel-3 OLCI and SLSTR instruments, and the NPP and JPSS VIIRS instrument successors to MODIS.

Gravity and the geoid

The measurement of small spatial and temporal variations in the Earth's gravity and of the geoid (i.e., equipotential reference surface) are required to quantify snow- and ice-mass distribution and mass transports including

freshwater exchange between land, ocean, ice, and atmosphere and for constraining heat, energy, and freshwater cycling through the Earth system.

Since 2000, three missions have contributed valuable data to meet these needs. The CHAMP, GRACE, and GOCE satellites have sequentially improved knowledge of the static gravity and geoid, while GRACE has revolutionized our understanding of the time-variable component of the gravity field. The GOCE mission promises to deliver a revolutionary accuracy in the static geoid over its operating lifetime. Gravity gradients from its fundamentally new gradiometer instrument, coupled with precise orbit information, give access to approximately 2 cm geoid accuracy at 100 km spatial resolution. This geoid promises a uniform global reference level for altimeter detection of regional dynamic ocean topography or sub-centimeter eustatic sea level changes (see entry [Sea Level Rise](#)) (over a nominal satellite lifetime of several years) in response to ice sheet melting (Wilson et al., 2010).

The GRACE follow-on (GRACE-FO) mission, currently being prepared for launch in 2016, is designed to provide a gap-free succession of data between the current GRACE mission and an upgraded GRACE-II planned for launch in the 2020 timeframe. GRACE-FO will continue to map the Earth's gravitational field with regional resolution and monthly variability.

Other infrastructure considerations

For the polar and cryospheric observing system to fulfill all essential needs, it is necessary to consider also a number of other critical issues which indirectly require other satellite remote sensing capabilities.

Ground-based observations

As opposed to remote sensing measurements, in situ measurements are made in direct contact with the medium of interest and typically result in sparse or sporadic observations in space or time (i.e., by comparison to the satellite data). Independent in situ, surface-based measurements of known quality of the physical state of the relevant cryospheric subdomain, or atmosphere, land, or ocean are required to understand the physical processes at work and to validate the satellite data products. Reference observation networks of autonomous or manned stations are needed for this purpose. Capabilities include precipitation (solid/liquid); snow water equivalent; vertical profiles of temperature, physical, and dielectric properties; satellite-tracked buoys (ice/ocean); automatic weather stations; and glacier or ice sheet GNSS survey reference sites.

Satellite communications and positioning

Operations in the polar regions and particular navigation in sea ice require a combination of satellite-based communications and accurate positioning. Many in situ observing measurement systems rely on some combination of

positioning, and satellite communications/data relay to operate successfully (e.g., autonomous drifting buoys).

Communications satellites are invaluable to telecommunications uses such as data relay in the polar regions. Modern communications satellites use a variety of orbits including geostationary orbits, elliptical (e.g., Molniya) orbits, and other and low (polar and nonpolar) Earth orbits. Communications satellites provide a microwave-based radio relay capability which today is used extensively for mobile applications such as communications to ships, aircraft, and handheld terminals, for which, in polar regions, alternative technologies do not exist. New highly elliptical orbiting polar communications and weather satellite systems such as the Canadian Polar Communications and Weather (PCW) and Russian Arktika satellites are currently in development to improve monitoring capability and high-latitude activities.

Today, accurate navigation and positioning has become dependent on satellite aids such as satellite navigation or Global Navigation Satellite Systems (GNSS) such as the currently operational US Global Positioning System (GPS) and the developing European GALILEO, the Russian GLONASS, and the Chinese BeiDou-2 (or COMPASS) systems.

Measurement reference frames

An accurate and stable measurement reference frame is essential for most precise remote sensing measurement techniques such as the altimetry and InSAR techniques discussed earlier (Wilson et al., 2010). This requires the following:

- Sustained support for satellite remote sensing tools integral to the International Terrestrial Reference Frame (ITRF), including satellite laser ranging, very-long-baseline interferometry, DORIS, and GNSS
- Inclusion of observations of the static gravity field from GOCE and other stand-alone missions to determine a precise, high spatial resolution geoid
- ITRF accurate to approximately 1 mm and stable to approximately 0.1 mm/year or better, in order to be able to understand the consequences of a changing cryosphere upon the rate of sea level change

Remote sensing product validation

One of the shortcomings in cryospheric remote sensing measurements is that algorithms used to retrieve geophysical data products may be sensitive to spatial variability, topography, atmospheric propagation, or other diurnally or seasonally driven physical effects. To check the robustness of measurements and veracity of the resulting products over time, it is necessary to perform product validation.

Independent validation measurement data are typically acquired during in situ experiments synchronized with the satellite overpass times, at specific carefully selected instrumented sites. In situ measurements must also be made at the appropriate time and space scale for

effective comparison with the satellite sampling. Comparison of well-calibrated in situ measurements with the satellite data products allows the evaluation of the consistency of measurements over time, and for the sources of uncertainties and biases in the satellite data to be rigorously quantified.

Validation of cryospheric products also relies on critical ancillary data provided by existing in situ measurement networks which report regularly via the Global Telecommunications System (GTS) or other satellite communication links. These networks include buoys in the ocean and sea ice pack, Global Observing System (GOS) weather stations and snow cover monitoring sites, and other terrestrial network sites comprising the Global Terrestrial Networks for Permafrost (GTN-P) and Glaciers (GTN-G).

Summary

Due to the complex interrelationships between the terrestrial, alpine, and marine elements of the cryosphere, the space component of the cryospheric and polar region observing system (*CryOS*) is needed to provide comprehensive information on all cryospheric domains. Methods of remotely sensed observations may be common to these domains, and thus the principal challenge for *CryOS* is to identify ways to develop, coordinate, maintain, and sustain these remote sensing observations within the GEOSS framework (Drinkwater et al., 2008; Jezek and Drinkwater, 2010).

Notably, the cryospheric and polar region observing system is recognized to require more than remote sensing measurements of snow and ice properties from satellites or airborne platforms. It must also include complementary networks of ground-based instrumentation as well as other capabilities such as modeling, data assimilation, and reanalysis systems and comprehensive data archiving and management systems.

Bibliography

- Alley, R. B., Clarke, P. U., Huybrechts, P., and Joughin, I., 2005. Ice-sheet and sea-level changes. *Science*, **310**(5747), 456–460.
- Breivik, L., Carrieres, T., Eastwood, S., Fleming, A., Girard-Ardhuin, F., Karvonen, J., Kwok, R., Meier, W., Mäkynen, M., Pedersen, L., Sandven, S., Similä, M., and Tonboe, R., 2010. Remote sensing of sea ice. In Hall, J., Harrison D. E., and Stammer, D. (eds.), *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*, Venice, September 21–25, 2009. ESA Publication, WPP-306, 2.
- Brown, J., Ferrians, Jr., O. J., Heginbottom, J. A., and Melnikov, E. S., 1997. *Circum-Arctic Map of Permafrost and Ground-Ice Conditions*. Washington, DC: U.S. Geological Survey in Cooperation with the Circum-Pacific Council for Energy and Mineral Resources. Circum-Pacific Map Series CP-45, scale 1:10,000,000, 1 sheet.
- Cavalieri, D., Parkinson, C., Gloersen, P., and Zwally, H. J., 1996. *Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data*. Boulder: National Snow and Ice Data Center. Digital Media.
- Comiso, J. C., Cavalieri, D. J., and Markus, T., 2003. Sea ice concentration, ice temperature, and snow depth using AMSR-E data.

- IEEE Transactions on Geoscience and Remote Sensing*, **41**(2), 243–252.
- Drinkwater, M. R., Jezek, K. C., and Key, J. R., 2008. Coordinated satellite observations during the international polar year 2007–2008: towards achieving a polar constellation. *Space Research Today*, **171**, 6–17.
- Drinkwater, M. R., Bonekamp, H., Bontempi, P., Chapron, B., Donlon, C., Fellous, J.-L., DiGiacomo, P., Harrison, E., LeTraon, P. Y., and Wilson, S., 2010. Status and outlook for the space component of an integrated ocean observing system. In Hall, J., Harrison, D. E., and Stammer, D. (eds.), *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*, Venice, Italy, September 21–25, 2009, ESA Publication, WPP-306, 1.
- Duguay, C. R., Zhang, T., Leverington, D. W., and Romanovsky, V. E., 2005. Remote sensing of permafrost and seasonally frozen ground. In Duguay, C. R., and Pietroniro, A. (eds.), *Remote Sensing in Northern Hydrology: Measuring Environmental Change*. Washington, DC: American Geophysical Union. Geophysical Monograph, Vol. 163, pp. 91–118.
- IGOS, 2007. *Integrated Global Observing Strategy Cryosphere Theme Report – For the Monitoring of our Environment from Space and from Earth*. Geneva: World Meteorological Organization., WMO/TD-No. 1405.
- Jeffries, M. O., Morris, K., and Kozlenko, N., 2005. Ice characteristics and processes, and remote sensing of frozen rivers and lakes. In Duguay, C. R., and Pietroniro, A. (eds.), *Remote Sensing in Northern Hydrology: Measuring Environmental Change*. Washington, DC: American Geophysical Union. Geophysical Monograph, Vol. 163, pp. 63–90.
- Jeffries, M. O., Morris, K., and Duguay, C. R., (in press). Lake ice and river ice. In Williams, Jr., R. S., and Ferrigno, J. G. (eds.), *Satellite Image Atlas of Glaciers of the World – Review of the State of the Earth's Cryosphere*. U.S. Geological Survey Professional Paper 1386-A.
- Jezek, K. J., and Drinkwater, M. R., 2010. Satellite observations from the international polar year. *Eos, Transactions American Geophysical Union*, **91**(14), 125–126.
- Kääb, A., 2008. Remote sensing of permafrost-related problems and hazards. *Permafrost and Periglacial Processes*, **136**, 107–136.
- Kwok, R., 2011. Satellite remote sensing of sea ice thickness and kinematics: a review. *Journal of Glaciology*, **56**(200), 1129–1140.
- Kwok, R., Cunningham, G. F., LaBelle-Hamer, N., Holt, B., and Rothrock, D. A., 1999. Ice thickness derived from high-resolution SAR imagery. *Eos, Transactions American Geophysical Union*, **80**(42), 495–497.
- Kwok, R., Cunningham, G. F., Wensnahan, M., Rigor, I., Zwally, H. J., and Yi, D., 2009. Thinning and volume loss of the Arctic Ocean sea ice cover: 2003–2008. *Journal of Geophysical Research*, **114**, C07005, doi:10.1029/2009JC005312.
- Lavergne, T., Eastwood, S., Teffah, Z., Schyberg, H., and Breivik, L.-A., 2010. Sea ice motion from low-resolution satellite sensors: an alternative method and its validation in the Arctic. *Journal of Geophysical Research*, **115**, C10032, doi:10.1029/2009JC005958.
- Long, D. G., Drinkwater, M. R., Holt, B., Saatchi, S., and Bertoia, C., 2001. Global ice and land climate studies using scatterometer image data. *Eos, Transactions American Geophysical Union*, **82**(43), 503.
- Perovich, D. K., Grenfell, T. C., Light, B., and Hobbs, P. V., 2002. Seasonal evolution of the albedo of multiyear Arctic sea ice. *Journal of Geophysical Research*, **107**(C10), 8044, doi:10.1029/2000JC000438.
- Rignot, E., Bamber, J. L., van den Broeke, M., Davis, C., Li, Y., van de Berg, W. J., and van Meijgaard, E., 2008. Recent Antarctic ice mass loss from radar interferometry and regional climate modeling. *Nature Geoscience*, **1**, 105–110.
- Rignot, E., Mouginot, J., and Scheuchl, B., 2011. Ice flow of the Antarctic ice sheet. *Science*, **333**(6048), 1427–1430.
- Van Den Broeke, M., Bamber, J., Ettema, J., Rignot, E., Schrama, E., van de Berg, W. J., van Meijgaard, E., Velicogna, I., and Wouters, B., 2009. Partitioning recent greenland mass loss. *Science*, **326**(5955), 984–986.
- Wilson, W. S., Abdalati, W., Alsdorf, D., Benveniste, J., Bonekamp, H., Cogley, J. G., Drinkwater, M. R., Fu, L. L., Gross, R., Haines, B. J., Harrison, D. E., Johnson, G. C., Johnson, M., LaBrecque, J. L., Lindstrom, E. J., Merrifield, M. A., Miller, L., Pavlis, E. C., Pietrovicz, S., Roemmich, D., Stammer, D., Thomas, R. H., Thouvenot, E., and Woodworth, P. L., 2010. Observing systems needed to address sea-level rise and variability. In Church, J. A., Woodworth, P. L., Aarup, T., and Wilson, W. S. (eds.), *Understanding Sea-Level Rise and Variability*. Oxford: Wiley-Blackwell, pp. 376–399.
- Zhang, T., Barry, R. G., Knowles, K., Ling, F., and Armstrong, R. L., 2003. Distribution of seasonally and perennially frozen ground in the Northern Hemisphere. In Phillips, M., Springman, S. M., and Arenson, L. U. (eds.), *Permafrost: Proceedings of the Eighth International Conference on Permafrost*, July 21–25, 2003, Zurich, Switzerland. Lisse, T Netherlands: A.A. Balkema.

Cross-references

[Calibration and Validation](#)
[Cryosphere, Climate Change Effects](#)
[Cryosphere, Climate Change Feedbacks](#)
[Cryosphere, Measurements and Applications](#)
[Geodesy](#)
[Global Climate Observing System](#)
[Global Earth Observation System of Systems \(GEOSS\)](#)
[Global Land Observing System](#)
[Ice Sheets and Ice Volume](#)
[Lidar Systems](#)
[Microwave Radiometers](#)
[Ocean Surface Topography](#)
[Polar Ice Dynamics](#)
[Radar, Altimeters](#)
[Radar, Scatterometers](#)
[Radar, Synthetic Aperture](#)
[Sea Ice Albedo](#)
[Sea Ice Concentration and Extent](#)
[Sea Surface Temperature](#)
[Sea Surface Wind/Stress Vector](#)
[Terrestrial Snow](#)

CRYOSPHERE, CLIMATE CHANGE EFFECTS

Aixue Hu
 Climate and Global Dynamics Division, National Center
 for Atmospheric Research, Boulder, CO, USA

Synonyms

Frozen water sphere; Solid water sphere

Definition

The cryosphere is defined as the part of the earth's surface where the water is in solid form. It includes the sea ice, lake and river ice, glaciers and mountain ice caps, ice sheets, snow cover, and the frozen ground (including permafrost). The cryosphere is important to global climate due to its high reflectivity (albedo) of the sun's rays. Changes in the cryosphere affect the global heat balance and thus modulate the global climate.

Albedo is the reflectivity of the earth's surface to the incoming solar radiation (sunlight).

Introduction

The cryosphere is an integral part of the earth's climate system. It is closely linked to the earth's surface energy budget, the water cycle, sea level changes, and the surface gas exchanges. Presently, the cryosphere on land contains about 75 % of the world's freshwater, mostly in the Greenland ice sheet and the Antarctic ice sheet. The total melt of these two ice sheets would increase the global sea level by about 64 m. In the past million years, the cryosphere has gone through significant changes on different time scales, such as the Ice Age. When the global climate cools, the cryosphere expands, and when the global climate warms, the cryosphere shrinks. On the other hand, the cryosphere not only responds passively to changes in the global climate, it also actively modifies the global climate through the albedo feedback. In general, roughly 80–90 % of the sun's rays reaching the snow or ice surface are reflected back into space. In contrast, only 10 % of the sun's rays are reflected back to space from the ocean surface. In a colder climate, the growth of sea ice and the snow cover and ice sheets on land would increase the albedo, which reflects more of the sun's rays back into space and reduces the heat absorption by the earth's climate system, leading to a further cooling effect of the climate. In a warmer climate, the shrinkage of the cryosphere would reduce the global albedo, thus allowing more of the sun's rays to be absorbed by the earth's climate system and make the climate even warmer. This is a positive feedback.

Present state of the cryosphere

Mean state of the cryosphere

Currently, about 10 % of the earth's land area is covered by ice, mostly in the Antarctic and Greenland; only a tiny fraction lies in the ice caps and mountain glaciers. In the annual mean, sea ice covers about 7 % of the earth's ocean area. In the Northern Hemisphere, the maximum sea ice extent is about 15 million square kilometers in winter, with a minimum of about 3.6 million square kilometers in summer in the Arctic and the adjacent seas. In the Southern Hemisphere, the sea ice covers about 19 million square kilometers of ocean in winter and about 3 million square kilometers in summer. In the Northern Hemisphere winter

season, snow covers about 33 % of the land north of the equator and reaches 49 % in midwinter. Since the land area south of 40 °S is very small in the Southern Hemisphere, the region with snow cover in winter is small, except for Antarctica. The frozen ground (including both seasonally frozen ground and permafrost) covers about 51 % of the land area in the Northern Hemisphere in winter, of which about half is permafrost.

Changes of the cryosphere during the twentieth century

Based on both historical and satellite observations, the snow cover has declined in spring and summer in the Northern Hemisphere since the early 1920s, especially since the late 1970s, but the changes are small in winter (Lemke et al., 2007). In the Southern Hemisphere, little or no long-term trend is found from the satellite observations for the last three decades, but there is substantial interannual variability.

For the river and lake ice, studies show that the freeze-up date in the late fall is later, with a rate of change of 5.8 ± 1.6 days per century, and the break-up date is earlier, changing at a rate of 6.5 ± 1.2 days per century for selected rivers and lakes in the Northern Hemisphere (Magnuson et al., 2000).

In terms of sea ice, the satellite observations show a significant declining trend (-7.4 ± 2.4 % per decade from 1979 to 2005) for the Northern Hemisphere summer sea ice extent (Comiso, 2003). The 2012 summer minimum sea ice extent was only about 3.6 million square kilometers, a dramatic reduction in comparison to the 1979–2000 average minimum of about seven million square kilometers. The winter sea ice cover is also shrinking in the Northern Hemisphere with a rate of about 44,000 km² per year since 1979. In the Southern Hemisphere, the sea ice cover increased slightly during the recent decade, but it is not statistically significant. On the other hand, from up-looking sonars on submarines and some moored instruments that measure sea ice thickness, a drastic thinning of the Arctic multiyear sea ice around the late 1980s is found (Johannessen et al., 2004). There is no trend before the late 1980s (Tucker et al., 2001).

The glaciers and ice caps which are not adjacent to the large ice sheets of Greenland and Antarctic occupy only a very small portion of the land surface. However, these glaciers and ice caps affect the river runoff since most of the large rivers originate there. Variations of the stability of these glaciers and ice caps would affect the river flow in spring and late summer. Studies show that the retreat of the glaciers and ice caps became evident globally as early as 1850 (Oerlemans, 2005). Regionally, the retreat of the glaciers and ice caps has been significant everywhere except in Europe and the Andes since 1960s (Dyurgerov and Meier, 2005). In Europe and the Andes, the glaciers and ice caps show evidence of retreat since the 1990s.

Since the Greenland and Antarctic ice sheets hold huge amounts of ice which could significantly raise the sea level if they were to fully melt, it is important to monitor the changes in the ice sheets and study the impact of these ice sheets on global climate in the future. In the Greenland ice sheets, about half of the ice loss is by surface melting and runoff into the sea. In the Antarctic, ice is drained by slow-moving ice at the center of the ice sheet and by faster-moving ice-walled ice streams or ice shelves and narrow ice tongues at the edges. Observations from multiple sources, including from remote sensing instruments, show that the Greenland ice sheet mass was balanced during most parts of the twentieth century, but the Greenland ice sheet started to lose mass since the early 1990s with an accelerating trend (Lemke et al., 2007). In the Antarctic, the ice sheet appears to be losing mass all the time which might be associated to the past forcing since portions of the ice sheet respond very slowly. Recent satellite observations indicate that there is a slight mass gain in the eastern Antarctic and a more significant mass loss in the west Antarctic (Rignot and Thomas, 2002; Zwally et al., 2006). The total contribution of the ice sheets, glaciers, and ice caps to the [sea level rise](#) has been about 1.2 ± 0.4 mm/year from 1993 to 2003.

Observations show a warming of the permafrost all over the world in the twentieth century, especially in the late twentieth century when data are more abundant (Lemke et al., 2007). In some regions, this warming causes the permafrost to degrade. The degradation of the permafrost can cause significant problems for human activities. Thawing can induce the subsidence of the ground, a downward displacement of the surface, especially for the ice-rich permafrost. Typically, this subsidence does not happen uniformly; some regions would rise up and some regions would sink. Any man-made infrastructure standing on the thawing permafrost can be significantly damaged. On the coast of the Arctic, the permafrost degradation causes an erosion of the coastline by up to 3 m per year. For the seasonally frozen ground, the active layer which thaws and freezes on top of the permafrost becomes thicker, indicating a downward propagation of the surface warming signal. The area of the seasonally frozen ground outside of the permafrost region has shrunk about 7 % in the Northern Hemisphere during the last century.

The cryosphere in past climates

In the past million years, the earth's climate has experienced a few cold-warm cycles – the glacial-interglacial cycle. Each of these cycles lasted about 100,000 years. The most recent interglacial started about 10,000 years ago. The cause of these glacial-interglacial cycles is not well understood, but strong evidence indicates that these are linked to regular variations of the earth's orbit around the sun, the so-called Milankovitch cycles. In each of these cycles, the amount of the solar radiation reaching the earth surface changes at each latitude in each season,

but may not change the annual mean value. Many studies suggest that the total amount of solar radiation reaching the North Hemisphere in summer is crucial for the starts and ends of this climate change cycle. If the snow in the previous winter can survive the following summer, snow cover would grow and ice sheets would grow due to the albedo feedback. A growing ice and snow cover on the earth's surface would reflect more sunlight into space and further cool the earth's climate, the ice sheets grow bigger, the glaciers and ice caps grow towards lower altitudes, and sea ice cover expands towards the equator. At the last glacial maximum (about 21,000 years before present), the global sea level dropped by about 120 m, or about 120 m of shallow ocean was turned onto land. The ice sheets covered a significant portion of the North America continent up to 45°N (called the Laurentide ice sheet), a large portion of Europe, and part of the western and middle Russia (called the Weichselian ice sheet) in the Northern Hemisphere and southern South America (called Patagonian ice sheet) and the Antarctic in the Southern Hemisphere. These ice sheets were, in general, about 3–4 km thick.

Projected changes of cryosphere in future climate

In the recent report of the Intergovernmental Panel on Climate Change (IPCC, 2007), scientists using comprehensive state-of-the-art coupled climate models show that the status of the sea ice cover, snow cover, and frozen ground in the future climate depends heavily on how much CO₂ is emitted into the atmosphere by human activities. The Arctic could be seasonally ice-free in the mid-to-late twenty-first century if the CO₂ emission were high (Meehl et al., 2007). If so, human and marine life would be affected dramatically. Polar bears rely on the sea ice to hunt for food. If there is no sea ice at all in summer, polar bears would not be able to obtain enough nourishment. Model projections also show a significant reduction of the snow cover in the Northern Hemisphere, a poleward movement of the permafrost extent (indicating a shrinking in permafrost area), and an increase in the active layer thickness in the permafrost region. These changes would reduce the surface albedo, and the earth system would absorb more solar radiation which would warm the earth's climate even more.

In recent years, a slight inland thickening and strong marginal thinning of the Greenland ice sheet produce a net loss of the ice mass. A model-projected warming in the twenty-first century would suggest a continued mass loss of the Greenland ice sheet. One model projection indicates that the entire Greenland ice sheet could completely melt away in the next 3,000 years if the current warm climate persists (Ridley et al., 2005). This would raise the global sea level by half a meter per century at peak melting. The changes in the Antarctic ice sheet as a whole are uncertain. But the Western Antarctic ice sheet is likely to lose more mass, or even totally collapse, if the future climate is warmer; a complete collapse of the West

Antarctic ice sheet would cause a global [sea level rise](#) of 5 m. It is hard to project how long this process will take from present ice sheet models.

In general, if the future climate is warmer than now projected by the IPCC models, it is certain the glaciers and ice caps will become smaller. Based on model simulations of 17 glaciers, the volume of loss of the glaciers is projected to be up to 60 % by 2050. Since the disappearance of these glaciers and ice caps is much faster than a potential reglaciation, the loss of the glaciers and ice caps may be irreversible, at least in some regions. Because the total mass of the global glaciers and ice caps is much smaller than the ice sheets of Greenland and Antarctic, the melting of these glaciers and ice caps would not raise the global sea level by much. However, it potentially can significantly impact the river runoff and the availability of freshwater supply, hence adversely affecting human activities.

Summary

The cryosphere is a very important part of the global climate system, and many important properties can be derived using remote sensing. The cryosphere played a significant role in the past changes of the earth's climate. Because of its high reflectivity (albedo) of solar radiation, shrinkage of the cryosphere coverage on the earth's surface would induce a higher absorption of the solar radiation by the earth's climate system which would enhance warming. A seasonal ice-free Arctic would enhance the hydrological cycle there, which will significantly affect human and marine life in the pan-Arctic region. A possible melt of the Greenland and West Antarctic ice sheets in the future would raise the global sea level markedly. Studies show that the climate was only a few degrees warmer during the last interglacial period than now and that the sea level was about 3–6 m higher. At that time, a very large portion of the Greenland ice sheet melted. If the sea level were to rise by 3 m in the next century or so, by the melting of the Greenland and West Antarctic ice sheets, human activities would be impacted in a momentous way.

Bibliography

- Climate change – Wikipedia, the free encyclopedia.
 Comiso, J. C., 2003. Large scale characteristics and variability of the global sea ice cover. In Thomas, D., and Dieckmann, G. S. (eds.), *Sea Ice – An Introduction to Its Physics, Biology, Chemistry, and Geology*. Oxford: Blackwell Science, pp. 112–142.
 Cryosphere – Wikipedia, the free encyclopedia.
 Dyurgerov, M., and Meier, M. F., 2005. Glaciers and the changing Earth system: a 2004 snapshot. *Occasional Paper 58, Institute of Arctic and Alpine Research*, University of Colorado, Boulder, CO, 118 pp.
 Global Outlook for Ice and Snow Assessment on the state and future of the Cryosphere, UN Environment Programme, June 2007.
 IPCC, 2007. Climate change 2007: the physical science basis. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (eds.), *Contribution of Working Group I to the Fourth Assessment Report of the*

- Intergovernmental Panel on Climate Change*. Cambridge/New York: Cambridge University Press, p. 996.
 Johannessen, O. M., et al., 2004. Arctic climate change – observed and modelled temperature and sea ice. *Tellus*, **56A**, 328–341.
 Lemke, P., et al., 2007. Observations: changes in snow, ice and frozen ground. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge/New York: Cambridge University Press, pp. 337–383.
 Magnuson, J. J., et al., 2000. Historical trends in lake and river ice cover in the Northern Hemisphere. *Science*, **289**, 1743–1746.
 Meehl, G. A., et al., 2007. Global climate projections. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge/New York: Cambridge University Press.
 Oerlemans, J., 2005. Extracting a climate signal from 169 glacier records. *Science*, **308**, 675–677.
 Ridley, J. K., Huybrechts, P., Gregory, J. M., and Lowe, J. A., 2005. Elimination of the Greenland ice sheet in a high CO₂ climate. *Journal of Climate*, **17**, 3409–3427.
 Rignot, E., and Thomas, R. H., 2002. Mass balance of polar ice sheets. *Science*, **297**(5586), 1502–1506.
 Tucker, W. B., III, et al., 2001. Evidence for the rapid thinning of sea ice in the western Arctic Ocean at the end of the 1980s. *Geophysical Research Letters*, **28**(14), 2851–2854.
 Zwally, H. J., et al., 2006. Mass changes of the Greenland and Antarctic ice sheets and shelves and contributions to sea level rise: 1992–2002. *Journal of Glaciology*, **51**, 509–527.

Cross-references

[Cryosphere, Climate Change Feedbacks](#)
[Cryosphere, Measurements and Applications](#)
[Climate Data Records](#)
[Climate Monitoring and Prediction](#)
[Ice Sheets and Ice Volume](#)
[Sea Ice Albedo](#)

CRYOSPHERE, CLIMATE CHANGE FEEDBACKS

Peter J. Minnett
 Rosenstiel School of Marine and Atmospheric Science,
 University of Miami, Miami, FL, USA

Definition

The term “feedback” is taken from electrical engineering and describes how a small part of the output of a circuit or a system, such as an amplifier, is fed back to become part of the input. A positive feedback loop results in a growth in the signal, or initial disturbance, and a negative feedback loop acts to diminish the effects of a disturbance, retaining the system close to a stable state. A positive feedback loop could push a hitherto stable system past a tipping point into another state. The concept of feedback loops is a very appealing idea for

geophysics and climate researchers, for example, as there are many levels of complexity that permit feedback loops of both signs that can stabilize the system to some disturbances or perturbations, and render it unstable to others (Bony et al., 2006).

Climate feedbacks

In the climate system, the simplest approach is to consider the change in the equilibrium temperature of the earth, or of a particular region (the temperature change being written as ΔT_d), that results directly from a change in the radiative forcing (ΔQ):

$$\Delta T_d = \lambda \Delta Q$$

where λ represents a sensitivity factor. For the direct radiative forcing of the climate system, λ is determined by the physical law relating radiation and temperature (Stefan-Boltzmann's Law) and has a value of about $0.25 \text{ K (Wm}^{-2})^{-1}$ for temperatures typical of the middle atmosphere. Thus, a change in radiative forcing of 4 Wm^{-2} would lead to a temperature change of about 1 K.

In the more realistic case, the climate changes involve feedbacks which modify ΔT_d to give the actual equilibrium temperature change (ΔT). Using the term "gain" (g) to represent the fraction of ΔT that is caused by feedback mechanisms, where g can be negative, but no larger than +1, then:

$$g = (\Delta T - \Delta T_d) / \Delta T$$

so that:

$$\Delta T = \Delta T_d / (1 - g)$$

Thus, a negative gain reduces the temperature change, and a value close to 1 leads to large amplification of the direct temperature change. There are generally several components in the climate gain factor which result from different physical mechanisms, and so the value of g is not a single value but varies according to the relative strengths of the different components, which in turn depend on the relative influence of the relevant parts of the climate system (NRC, 2003).

In the cryosphere, several feedbacks are also present in other parts of the climate system, but here they are rendered more extreme by the special characteristics of a region that are close to the freezing point of water. Several of the various components of the cryospheric feedbacks are accessible to satellite remote sensing, either directly through measurements of the critical variables, or through measurements of the consequences of the feedbacks.

In the polar regions, there are at least four major feedbacks that can, at least in principle, be monitored, or quantified, using remotely sensed data: the surface albedo-shortwave radiation feedback, the water-vapor feedback, the cloud radiative feedbacks, and the biogenic cloud condensation nuclei (CCN) feedback.

Surface albedo-shortwave radiative feedback

The rapid loss of summertime Arctic Ocean ice cover in recent years has focused attention on possible mechanisms for the accelerated loss, and the surface albedo feedback is one that is frequently mentioned. When the ocean is covered by ice, most of the incident solar radiation is reflected back through the atmosphere and to space. The ratio of the total amount of reflected energy to the incoming sunlight is called the albedo of the surface. Dry, fresh snow on sea ice can reflect as much as 90 % of the incident sunlight. The reflectivity varies with wavelength of the radiation (Hanesiak et al., 2001). As the temperature rises and the snow or ice begins to moisten, the albedo decreases, resulting in more of the solar energy being absorbed, and this leads to more melting of snow and ice. When melt ponds form, the surface albedo falls further and more solar energy is absorbed, leading to accelerated melting. The melting sea ice reveals the dark sea surface which has an albedo of <10 %. The heat resulting from the absorption of solar energy can be moved laterally by surface currents and brought into contact with sea ice, which it can melt from the sides or from beneath.

The situation on land is similar: when snow and ice melt away, they are replaced by bare ground or vegetation which generally has lower albedo and therefore absorbs more energy, hastening the melting of the remaining snow and ice nearby.

The albedo shortwave radiation feedback is a positive one, as the melting of the snow and ice reduces the surface albedo, resulting in the absorption of more sunlight, leading to yet more snow and ice melt (Curry et al., 1995).

The extent of polar sea ice is monitored very effectively from polar orbiting satellites using microwave radiometers and the surface reflectivity, and hence albedo is measured by satellite imaging radiometers operating in the visible part of the electromagnetic spectrum.

Water-vapor feedback

Water vapor in the atmosphere is a very potent greenhouse gas. This means the water molecules at a given height in the atmosphere absorb some of the infrared radiation emitted from the surface below and from the intervening atmosphere, radiation that would otherwise escape to space thereby cooling the planet; when that energy is reemitted, some of it propagates back toward the surface. The maximum amount of water vapor the atmosphere can hold without condensing is approximately exponentially dependent on the absolute temperature. Thus, as the temperature of the air rises, so does its ability to support an increasing water vapor burden and therefore its ability to intercept more of the outgoing infrared radiation. This is a positive feedback. The actual amount of water vapor in the atmosphere is also dependent on an available source of water, usually at the base of the atmosphere. A frozen surface is a poor source of moisture, so the atmosphere above is generally very dry, and the downward infrared radiation from the

atmospheric water vapor emission is small. Over open water, evaporation into the atmosphere leads to a moister atmosphere and increased downward infrared radiative heating of the surface. The advection of the moister atmosphere from over the open water to over sea-ice increases the radiative heating of the surface, leading to melting of the surface and increased evaporation – another positive feedback.

Although the atmospheric water vapor content, often called the precipitable water, can be measured quite accurately over the open oceans by microwave radiometers on spacecraft, the retrievals are not very accurate over land or over sea ice, or a surface comprised of mixed open water and ice. The atmospheric water vapor profile can be derived from measurements of infrared and microwave sounding radiometers, but again over land and over mixed ice and water surface, the retrievals are not very accurate. Thus the remote sensing of the water vapor and hence the direct quantification of this feedback is not possible using current remote sensing technology.

Cloud radiative feedbacks

A consequence of the increased atmospheric water vapor content resulting from evaporation over open water in polar regions is increased cloud cover. Often the low-level clouds that form over leads and polynyas (areas of open water surrounded by ice; Smith and Barber, 2007) can be traced many kilometers downwind, and are apparent in satellite imagery (Dethleff, 1994). Over the polar regions, as elsewhere, clouds can be carried great distances by the winds. The presence of clouds has two consequences on the surface radiative energy budget: they heat the surface by increasing the infrared radiation incident at the surface, and they cool the surface by casting shadows, that is, by scattering sunlight back to space. The former is a positive feedback, and the latter a negative one (Curry et al., 1996). Both are present at the same time, and the combined effect depends on the relative strengths of each (Hanafin and Minnett, 2001). The infrared heating depends on the amount, type, thickness, and height of the clouds and is present both night and day. The short-wave cooling occurs only in daylight and depends on the amount, type of clouds, and also on the solar zenith angle, and whether there are clouds across the face of the sun (Minnett, 1999). In general, the negative feedback that results in the cooling of the surface occurs in the polar summer, when the sun is higher in the sky, whereas the positive feedbacks occur when the incident solar radiation is small, when the sun is close to the horizon, or absent, as in the winter (Intrieri et al., 2002; Vavrus, 2004).

Clouds can be identified in satellite images, although this is more readily done when the clouds are illuminated by sunlight against a dark background of open water or dark vegetation. Over a bright frozen surface, the automatic identification of clouds is difficult and appreciable errors can occur. Similarly at night, when the

identification of clouds relies on a temperature contrast with respect to the surface, problems of misidentification can arise. Relatively new techniques using space-borne cloud radars and lidars show promise in improving the confident identification and classification of clouds and consequent improvements in the study of the cloud radiative feedbacks in the polar regions.

Biogenic cloud condensation nuclei (CCN) feedback

For clouds to form in the polar atmosphere, it is not sufficient that the water vapor content be at or above saturation, but also that there are small particles in the air that can act as cloud condensation nuclei. One group of CCNs over the oceans are minute salt crystals that result from the evaporation of spray droplets and another are derived from biological activity in the ocean. Airborne experiments over the Antarctic ice floes revealed bacteria and algal spores acting as CCNs (Saxena, 1983), and subsequently dimethylsulfide (DMS), a gas released from its precursor dimethyl sulfonium propionate (DMSP) at the ocean surface, have been found to be effective CCNs. DMSP is released when phytoplankton and algae die and decay. More recently, viruses, bacteria, and fragments of diatoms have been found to act as CCNs in the lower atmosphere over leads in the Arctic Ocean (Leck et al., 2004). The presence of biogenic material provides a possible feedback mechanism in the climate system: solar radiation is necessary for plankton to bloom, and when they die they release material that become CCNs, increasing the cloud amount, and this leads to a reduction in solar radiation reaching the surface – a negative feedback.

The presence of CCNs promotes not only cloud formation, but also influences the cloud properties. High concentrations of CCNs result in more and smaller droplets in the clouds, and this has two consequences. Firstly, the clouds have a higher albedo and therefore scatter more solar energy back to space, the Twomey Effect (initially described in terms of anthropogenic pollution in the atmosphere (Twomey, 1977)), and secondly, the smaller cloud droplets take longer to accrete to sizes that will lead to precipitation, so that the clouds live longer – the Albrecht Effect (Albrecht, 1989).

This biogenic negative feedback has been hypothesized to result in the appropriate amount of sunlight reaching the ocean surface to sustain the appropriate levels of biological activity. Too little sunlight reaching the surface, because of a local positive perturbation in the cloud cover, leads to less phytoplankton growth, fewer CCNs, and eventually more sunlight reaching the surface because fewer clouds are formed, or they have properties that are less effective in reducing the sunlight at the surface (Leck et al., 2004).

Unlike the other feedbacks, discussed briefly here, the biogenic feedback remains hypothetical and the links involved have not been quantitatively

substantiated. It remains a topic of research. The chlorophyll concentration in the upper ocean can be monitored using remotely sensed measurements of ocean color derived from reflected sunlight in the visible part of the spectrum, and clouds can be monitored from space, as mentioned above.

Summary

Climate feedbacks in polar regions act to accelerate or delay the effects of global change, not only at high latitudes, but worldwide. Because of the remoteness of the polar regions, and the difficulty of gathering appropriate measurements using conventional instrumentation, the study and monitoring of many of the components of the feedback loops are feasible only by using satellite remote sensing.

Bibliography

- Albrecht, B. A., 1989. Aerosols, cloud microphysics and fractional cloudiness. *Science*, **245**, 1227–1230.
- Bony, S., Colman, R., Kattsov, V. M., Allan, R. P., Bretherton, C. S., Dufresne, J.-L., Hall, A., Hallegatte, S., Holland, M. M., Ingram, W., Randall, D. A., Soden, B. J., Tselioudis, G., and Webb, M. J., 2006. How well do we understand and evaluate climate change feedback processes? *Journal of Climate*, **19**, 3445–3482.
- Curry, J. A., Schramm, J. L., and Ebert, E. E., 1995. Sea ice-albedo climate feedback mechanism. *Journal of Climate*, **8**, 240–247.
- Curry, J. A., Rossow, W. B., Randall, D., and Schram, J., 1996. Overview of Arctic cloud and radiation characteristics. *Journal of Climate*, **9**, 1731–1764.
- Dethleff, D., 1994. Polynyas as a possible source for enigmatic Bennett Island atmospheric plumes. In Johannessen, O. M., Muench, R. D., and Overland, J. E. (eds.), *Polar Oceans and their Role in Shaping the Global Environment*. Washington, DC: American Geophysical Union, Vol. 85, pp. 475–483.
- Hanafin, J. A., and Minnett, P. J., 2001. Cloud forcing of surface radiation in the North Water Polynya. *Atmosphere-Ocean*, **39**, 239–255.
- Hanesiak, J. M., Barber, D. G., deAbreu, R. A., and Yackel, J. J., 2001. Local and regional albedo observations of arctic first-year sea ice during melt ponding. *Journal of Geophysical Research*, **106**, 1005–1016.
- Intrieri, J. M., Fairall, C. W., Shupe, M. D., Persson, P. O. G., Andreas, E. L., Guest, P. S., and Moritz, R. E., 2002. An annual cycle of Arctic surface cloud forcing at SHEBA. *Journal of Geophysical Research*, **107**, 8039, doi:10.1029/2000JC000439.
- Leck, C., Tjernström, M., Matrai, P., Swietlicki, E., and Bigg, K., 2004. Can marine micro-organisms influence melting of the Arctic pack ice? *Eos Transactions, American Geophysical Union*, **85**, 25.
- Minnett, P. J., 1999. The influence of solar zenith angle and cloud type on cloud radiative forcing at the surface in the Arctic. *Journal of Climate*, **12**, 147–158.
- NRC, 2003. *Understanding Climate Change Feedbacks*. Washington, DC: National Academy of Sciences.
- Saxena, V. K., 1983. Evidence of the biogenic nuclei involvement in Antarctic coastal clouds. *Journal of Physical Chemistry*, **87**, 4130–4134.
- Smith, W. O., and Barber, D. G. (eds.), 2007. *Polynyas: Windows to the World*. Amsterdam: Elsevier. Elsevier Oceanography Series.

- Twomey, S. A., 1977. The influence of pollution on the shortwave albedo of clouds. *Journal of the Atmospheric Sciences*, **34**, 1149–1152.
- Vavrus, S., 2004. The impact of cloud feedbacks on arctic climate under greenhouse forcing. *Journal of Climate*, **17**, 603–615.

Cross-references

Climate Monitoring and Prediction
 Cloud Properties
 Cryosphere, Climate Change Effects
 Cryosphere, Measurements and Applications
 Emerging Technologies, Lidar
 Ice Sheets and Ice Volume
 Polar Ice Dynamics
 Sea Ice Albedo
 Sea Ice Concentration and Extent
 Terrestrial Snow

CRYOSPHERE, MEASUREMENTS AND APPLICATIONS

Roger Barry
 National Snow and Ice Data Center, NSIDC 449 UCB,
 University of Colorado, Boulder, CO, USA

Definitions

(Land)fast ice. Sea ice attached to the shore or near-shore sea bed.

Nunatak. A mountain peak protruding through glacier ice.

Introduction

Remote sensing of the cryosphere has a 50 year history. We begin with an historical overview of aerial photography of the cryosphere – mainly glaciers and ice caps – and then proceed to satellite imagery and data for each of the major components of the cryosphere. In turn, we consider glaciers, snow cover, freshwater ice, sea ice, ice sheets and ice shelves, icebergs, and frozen ground.

Aerial photography

Glaciers and ice caps

Remote sensing of the cryosphere began with both vertical and oblique aerial photography of glaciers and ice caps. A massive campaign to photograph the Canadian Arctic was undertaken in the late 1940s–1950s (Dunbar and Greenway, 1956) and the photography covered all land areas. The US Geological Survey (USGS) photographed numerous glaciers in western North America and Alaska beginning in the 1950s (Post, 2005: <http://earthweb.ess.washington.edu/EPIC/Collections/Post/index.htm>).

The Austin Post aerial photograph collection begins in 1957 and continued through the 1980s. About 100,000 images on microfilm became a part of the World Data Center (WDC) for Glaciology collection when the WDC was transferred to the USGS in Tacoma in 1970 and to

the University of Colorado, Boulder, in 1976. This collection of the USGS Ice and Climate Project is known as the Post-Mayo-Krimmel Collection.

Aerial photography has continued to find applications. For example, snow patches remaining in late July in the Canadian Arctic have been mapped from air photographs for 1948–1983 by Lauriol et al. (1986). They found a high (0.93) correlation between the residual snow and maximum late winter snow thickness.

Satellite imagery and data

Aerial photography studies largely transitioned to satellite imaging from the Earth Resources Technology Satellite (ERTS) in July 1972 (renamed Landsat 1) and the subsequent Landsat series through Landsat 7 (15 m resolution) launched in 1999. The images are from a return beam vidicon (RBV) camera and multispectral scanner (MSS).

US military satellites in the CORONA and related programs collected over 800,000 diffraction-limited panoramic photographs with 2–7 m ground resolution from 1960 to 1972, which were declassified and made available for scientific use in the 1990s (McDonald, 1995). Mosaics of Greenland were developed from the Declassified Intelligence Satellite Photographs (DISP) by Zhou and Jezek (2002) in 1962 and 1963. Csatho et al. (2000) discuss the rectification of CORONA images for studying Greenland glacier motion and Altmaler and Kany (2002) use stereographic pairs for digital surface model generation.

In March 2002, the European Space Agency launched Envisat with the MEdium Resolution Imaging Spectrometer (MERIS) which has 15 channels in the 390–1,040 nm spectral range; a spatial resolution at nadir of 300 m is reduced on board to 1,200 m. In January 2006, the Japan Aerospace Exploration Agency (JAXA) launched the Advanced Land Observing Satellite (ALOS) with three remote sensing instruments: the Panchromatic Remote sensing Instrument for Stereo Mapping (PRISM) for digital elevation mapping with 2.5 m spatial resolution, the Advanced Visible and Near-Infrared Radiometer type 2 (AVNIR-2) for precise land coverage observation, and the Phased Array type L-band Synthetic Aperture Radar (PALSAR) for day-and-night and all-weather land observation. Imagery from commercial satellites has found occasional use but is too expensive for most research purposes. This includes data from the French Satellite Pour l'Observation de la Terre (SPOT) series, with SPOT 1 launched in 1986 and 5 in 2002 with 2.5–20 m resolution, the US Ikonos launched in 1999 with 1–4 m resolution and Quickbird launched in 2001 with 0.6–2.4 m resolution. The status of Russian environmental satellites in the METEOR, OKEAN, and RESURS series is reported by Asmus (2003).

Glaciers

Glacier mapping from Landsat Thematic Mapper (TM) data has followed several approaches (Paul, 2000). These involve: (1) manually delineating the glacier outline by

cursor tracking, (2) segmentation of ratio images, and (3) unsupervised or supervised classification. The first approach has been used to determine glacier length changes (Hall et al., 1992, for example). Various combinations of ratios have been used. Bayr et al. (1994) derive a glacier mask by using thresholds with ratio images of raw digital numbers from TM channels 4 and 5. Rott and Markl (1989) use atmospherically corrected spectral images of TM3/TM5 and TM4/TM5, with the help of thresholds to obtain a glacier mask. Paul noted that classification methods, both supervised and unsupervised, proved not to be suitable for glacier mapping.

Beginning in 1978, a major effort was undertaken by Williams and Ferrigno (1988) of the US Geological Survey to compile a Satellite Image Atlas of Glaciers of the World. The authors used maps; aerial photographs; Landsat 1, 2, and 3 MSS images; and Landsat 2 and 3 RBV images to inventory the areal distribution of glacier ice between about 82° north and south latitudes. Some later contributors also used Landsat 4 and 5 MSS and Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper-Plus (ETM+), and other satellite images. There are 10 regional chapters: Those concerning Antarctica (B); Greenland (C); Continental Europe (E); Asia (F); Turkey, Iran, and Africa (G); Irian Jaya, Indonesia, and New Zealand (H); South America (I); North America excluding Alaska (J); and Alaska (K). All chapters are now accessible at <http://pubs.usgs.gov/pp/p1386/>.

A NASA-funded project for Global Land Ice Measurements from Space (GLIMS) is under way at the National Snow and Ice Data Center (NSIDC) (Armstrong et al., 2005b; <http://www.glims.org/>). It is an international project involving 60 institutions worldwide, with the goal of surveying a majority of the world's estimated 103,000 glaciers (Bishop et al., 2004). Currently over 96,000 glacier outlines are in the GLIMS database. A supplemental globally complete inventory was developed for the IPCC AR5 report and is available as the Randolph Glacier Inventory RGI3.0 (April 2013) from GLIMS. GLIMS uses data collected primarily by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) instrument, aboard the Terra satellite, and the LANDSAT ETM+, along with historical observations. Analysis of changes in glacier extent using GLIMS data, earlier satellite imagery, and historical maps for the Eastern Pamir (Khromova et al., 2006) show that the glacier area decreased 7.8 % during 1978–1990, and 11.6 % in 1990–2001. In the Cordillera Blanca, Peru, Raup et al. (2007) show that glaciers receded by 11–30 % over the period 1962–2003; they also recognized inconsistencies in the earlier glacier mapping. Racoviteanu et al. (2008) use Spot 5 scenes from 2003 compared with aerial photography from 1970 for the Cordillera Blanca and report a 22.4 % loss in glacier area. High-resolution ALOS/PRISM data have been used for mapping glaciers in Eurasia and determining areal changes by comparison with earlier data by Aizen et al. (2007) and Surazakov et al. (2007).

There is still no single map of the world's glaciers. However, work is under way at the NSIDC to compile such a map using 500 m-resolution Moderate-Resolution Imaging Spectroradiometer (MODIS) images. The procedure involves determining whether a pixel has ever been snow free, by combining the lowest reflectance signatures for each pixel over several summer time periods in order to identify permanent snow and ice with the highest reflectance.

Glacier motion has been mapped with JERS-1 L-band SAR data. Strozzi et al. (2008) employ offset fields between pairs of JERS-1 satellite SAR data acquired in winter with 44 days time interval to estimate glacier motion on Svalbard, Novaya Zemlya, and Franz-Josef Land. The displacement maps show that the ice caps are divided into a number of clearly defined fast-flowing units with displacement larger than about 50 m/year with an estimated error on the order of 20 m/year.

Glacier mass balance has been estimated for three glaciers in the French Alps using a combination of SAR data and surface stakes (Dedieu et al., 2003). The essential key to using remote sensing is to obtain a high-resolution digital elevation model (DEM). This is possible from ASTER using stereoscopy, provided there are ground control points (GCPs) on and around the glaciers.

Snow cover

Hemispheric analysis of snow cover began in the United States in October 1966 from the polar orbiting Very High Resolution Radiometer (VHRR), with 3 km resolution, and continued with the use of AVHRR, GOES, and other mainly visible-band satellite data. The National Oceanic Atmospheric Administration (NOAA) issued weekly snow cover maps of the Northern Hemisphere that, beginning in 1972, were analyzed from the Advanced Very High Resolution Radiometer (AVHRR) instrument with 1.1 km resolution. The NOAA snow maps were generated from a polar stereographic grid whose pixel size ranged from 125×125 – 200×200 km depending on latitude. This now provides the longest time series of any satellite product. Since February 1997, it has been issued on a daily basis using the Interactive Multisensor Snow and Ice Mapping System (IMS) (Ramsay, 1998). Recent upgrades to the IMS are reported by Helfrich et al. (2007). They include: a 4 km-resolution grid output, ingest of an automated snow detection algorithm, expansion to global extent, and a static Digital Elevation Model for mapping based on elevation.

Global snow cover maps are now available from MODIS on Terra (February 2000–present) and Aqua (July 2002–present). There are daily, 8 day, and monthly global grids at 500 m and 0.05° resolution (<http://nsidc.org/data/modis/data.html>). Hall and Riggs (2007) provide an accuracy assessment for the Terra products. The overall absolute accuracy of the 500 m resolution swath and daily tile products is 93 %, but varies by land-cover type and

snow condition. The most frequent errors are due to snow/cloud discrimination problems; however, improvements in the MODIS cloud mask have occurred in each reprocessing. Detection of very thin snow (<1 cm thick) can also be problematic.

Global snow depth and water equivalent can be estimated from passive microwave data. One of the earliest algorithm applications was by Foster et al. (1980) using SMMR data for a hemispheric analysis. Chang et al. (1987, 1990) use the difference of the 18 and 37 GHz channels on SMMR to estimate snow volume in the Northern Hemisphere. The mean monthly snow volume ranges from about 1.5×10^{13} kg in summer to about 300×10^{13} kg in winter. Tait (1998) developed a series of algorithms for different land-cover types similar to those of Chang et al. (1987). Snow grain size, especially the presence of large depth hoar crystals, has a huge effect on microwave backscatter and so decreases the brightness temperature. Over the Canadian Prairies, operational analyses have been issued in near-real time since 1988/1989 using a vertically polarized gradient ratio algorithm $(37-19 \text{ V})/18$ (Goodison and Walker (1994). The algorithm was much improved by the use of a wet snow indicator. Foster et al. (1997) compare snow mass estimates from a prototype passive microwave snow algorithm, a revised algorithm, and a snow depth climatology.

Global maps of snow water equivalent from SMMR and SSM/I have been produced from 1978 to 2007 (with updates) by Armstrong et al. (2005a); (<http://nsidc.org/data/nsidc-0271.html>). The data are in 25 km grids but the radiometric information is from an area that is larger than 625 km^2 , so the gridded value represents a mean Snow Water Equivalent (SWE) for this area. There is decreased confidence in the SWE reliability and possible under-measurement in the following cases:

1. Mountainous areas with large topographic variability return low SWE values. Samples from these areas contain a mixed signal from a large footprint.
2. Forested areas return low mean SWE values, because the mixed signal includes emission from trees and the snow canopy as well as the underlying surface.
3. Areas near coastlines return low or no SWE values, because the mixed signal includes frozen and unfrozen water and land.
4. Areas containing melting snow or wet snow packs typical of maritime snow conditions return low or no SWE values, because the microwave emission from liquid water overwhelms scattering from the snow pack.
5. Shallow or intermittent snow during fall and early winter typically does not result in sufficient microwave scattering to reliably detect SWE.

Further details are available at http://nsidc.org/data/docs/daac/nsidc0271_ease_grid_swe_climatology.gd.html

The Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) was launched on Aqua in May 2002. It measured horizontally and vertically polarized brightness temperatures at 6.9, 10.7, 18.7, 23.8,

36.5, and 89.0 GHz. The spatial resolution of the individual measurements varied from 5.4 km at 89 GHz to 56 km at 6.9 GHz. Thus, it had twice the resolution of SSM/I. A snow cover algorithm has been developed and tested by Chang et al. (2003). AMSR-E (SWE) and MODIS (snow extent) are available at NSIDC. AMSR-E was decommissioned on October 4, 2011.

Regional analyses of snow cover extent and characteristics have been carried out using Landsat Thematic Mapper data (Dozier and Marks, 1987; Martinec and Rango, 1987).

Subsequently, Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data have been used to map grain size and other snow characteristics via spectral mixture analysis. AVIRIS has 224 contiguous spectral channels with wavelengths from 400 to 2,500 nm. The spatial resolution is 4–20 m. The classification algorithm is based on the multiple end-member approach to spectral mixture analysis in which the spectral endmembers and the number of endmembers can vary on a pixel-by-pixel basis. This approach accounts for surface cover heterogeneity within a scene. The mixture analysis is operated on endmembers from a library of snow, vegetation, rock, soil, and lake ice spectra. Snow end members of varying grain size were produced with a radiative-transfer model (Painter et al., 1998, 2003; Dozier and Painter, 2004).

Algorithms are available to map snow extent (Shi and Dozier, 1993; Rott and Nagler, 1993; Shi and Dozier, 1997) and snow depth and density (Shi and Dozier, 1996) using SAR data. These approaches are valuable for small areas in mountain terrain but the limited swath coverage makes them impractical for continental regions.

The Cold Land Processes Experiment (CLPX), funded by NASA, was a multisensor, multi-scale approach to assess terrestrial snow cover, providing a comprehensive data set necessary to address several experiment objectives. Within a framework of nested study areas in the Rocky Mountains of Colorado, ranging from 1 ha to 160,000 km², intensive ground, airborne, and spaceborne observations were collected for mid-winter and spring intensive observation periods (IOPs) in 2002–2003. The three major questions of the mission (Cline et al., 1999) were as follows:

1. Can the components of the terrestrial cryosphere be observed from space accurately enough to identify meaningful climatic trends?
2. To what extent can snow and frozen ground information, deduced from remote sensing data, improve models of cold season processes, hydrologic forecasts, and forecasts of high-latitude ecosystem functions?
3. To what accuracy can SWE be estimated from remote sensing data, and is this sufficient for hydrological applications?

More specifically, the project aimed to:

- Evaluate and improve snow water equivalent retrieval algorithms for spaceborne passive microwave sensors such as SSM/I and AMSR-E

- Evaluate and improve radar retrieval algorithms for snow depth, density, and wetness, and soil freeze/thaw status
- Improve radar retrieval algorithms to enable discrimination of freeze/thaw status of snow, soil, and vegetation surfaces
- Examine the effects of spatial resolution on the skill of active and passive microwave remote sensing retrieval algorithms for snow and freeze/thaw status
- Examine the feasibility of coupling forward microwave radiative-transfer schemes to spatially distributed snow/soil models, to improve assimilation of microwave remote sensing data
- Examine the spatial variability of snow and frozen soil distributions in different environments and (a) improve the representation of subgrid-scale variability of snow and frozen soil in coupled and uncoupled land surface models, and (b) improve the representation of orographic precipitation (snowfall) in atmospheric models

Data collection focused on airborne measurements from five sensors each with a different spatial resolution. The NASA AIRSAR instrument collected synthetic aperture radar measurements at three frequencies (P-, L-, and C-bands) in both polarimetric and interferometric modes. The NASA POLSCAT instrument collected Ku-band scatterometer measurements. The NOAA PSR-A instrument collected passive microwave measurements at five frequencies ranging from 10.7 to 89 GHz. The similar Airborne Earth Science Microwave Imaging Radiometer (AESMIR) was flown during the 2003 campaigns to collect passive microwave measurements (at all AMSR-E frequencies from 6.9 to 89 GHz). The National Weather Service (NWS)/National Operational Hydrologic Remote Sensing Center (NOHRSC) airborne snow survey program also flew similar snow-detection sensors. An instrument measured terrestrial and atmospheric gamma radiation, which was used to determine snow water equivalent based on standard operational algorithms. At the two largest scales (33,000 and 142,000 km²), data collection focused on spaceborne measurements from several active and passive microwave and optical sensors.

Tedesco et al. (2005) show that airborne Polarimetric Scanning Radiometer data can be modeled by a log-normal distribution (Fraser, forested area) and by a bimodal distribution (North Park, patchy-snow, non-forested area). The brightness temperatures are resampled over a range of resolutions to study the effects of sensor resolution on the shape of the distribution, on the values of the average brightness temperatures, and on the standard deviations. The histograms become more uniform and the spatial information contained in the initial distribution is lost for a resolution >5,000 m in both areas. Tedesco et al. (2006) analyze the brightness temperatures of melting and refreezing snow using a truck-mounted radiometer at 6.7, 19, and 37 GHz and measuring brightness as a function of snow wetness. The physical model

reproduces brightness temperatures with a relative error of 3 % (8 K).

Haran (2003) provides MODIS radiances, reflectances, snow cover, and related grids for the CLPX IOPs in 2002–2003. Stankov and Gasiewski (2004) supply airborne multiband polarimetric brightness temperature images over three 25×25 km meso-scale study areas for both years. A listing of CLPX data sets is available at: <http://www.nsidc.org/data/clpx/#data>.

Freshwater ice

Ice forms seasonally on lakes, reservoirs, and rivers. A major problem for detection is the small spatial scale of rivers and small lakes. Detection of ice cover by passive microwave data is possible on large lakes and ice conditions have been routinely mapped on Great Slave and Great Bear Lakes in Northern Canada since April 1992 (Walker and Davey, 1993). DMSP SSM/I data are acquired with a focus on ice freeze-up and breakup. It has been found possible to discriminate between areas of ice cover and open water using SSM/I 85 GHz data. Airborne SAR data were used by Leconte and Klassen (1991) to determine lake and river ice features in Northern Manitoba, and Radarsat SAR is used routinely for ice classification and mapping over the Great Lakes (Leshkevich et al., 1998). Wynne et al. (1998) determined lake ice breakup dates from 1980 to 1994 for 81 lakes and reservoirs in the US Upper Midwest and portions of Canada (60°N , 105°W to 40°N , 85°W). Analyses of images from the visible band of the GOES-VISSR were used. The objectives were to investigate the utility of monitoring ice phenology as a climate indicator and to assess regional trends in lake ice breakup dates. MODIS 250 m-resolution data have been used for a study of ice-out dates in the Brooks Range (Kukthuroja et al., 2006). An initial problem was to construct a lake mask using 30 m TM imagery. Duguay and Lafleur (2003) combine optical and SAR data analysis to map ice thickness on shallow sub-Arctic lakes.

Sea ice

Sea ice in the Eurasian Arctic seas was routinely mapped by visual reconnaissance from aircraft flights in the Soviet Union starting in July 1933 and continuing until 1992 (Borodachev and Shilnikov, 2003). From the 1950s onward, 30–40 aircraft made 500–700 flights annually (Johannessen et al., 2007). The coverage was initially only in late summer, but by 1950, it was continuous throughout the year. Side-Looking Airborne Radar (SLAR) mapping was used from the mid-1960s and in 1983. SLR (Side-looking RADAR) was available from the Okean 01 series of satellites. Ice concentration and ice type were mapped at 10 to 30 day intervals. The chart records have been resumed since 1997 using satellite data. Early paper charts were digitized, and the entire series, including later charts that were produced entirely digitally, were converted to Sea Ice Grid (SIGRID) format at the Arctic and Antarctic Research Institute (AARI) in St. Petersburg, Russia.

NSIDC completed the conversion to EASE-Grid – a Lambert equal-area projection with 12.5 km cell size. Total ice concentration, as well as partial concentrations for multiyear, first-year, new/young ice, and fast ice, is available (Arctic and Antarctic Research Institute, 2007). Mahoney et al. (2008) analyzed sea ice extent in the Eurasian Arctic for 1933 to 2006 using the Soviet historical data.

The first aerial ice reconnaissance in Canada was completed during the winter of 1927–1928 by the Royal Canadian Air Force (RCAF) over Hudson Strait and Hudson Bay. In 1940, the Canadian Department of Transport Marine Services began an “Ice Patrol” in the Gulf of St. Lawrence. Summer patrols in the Arctic began in 1957. The first SLAR used for ice reconnaissance was installed in 1978; it had a 100 m resolution. SLAR measurements continue to be used along the eastern coast of Canada. Airborne Synthetic Aperture Radar (SAR) was introduced in 1990 with digital processing techniques and a resolution in the range of 5–30 m.

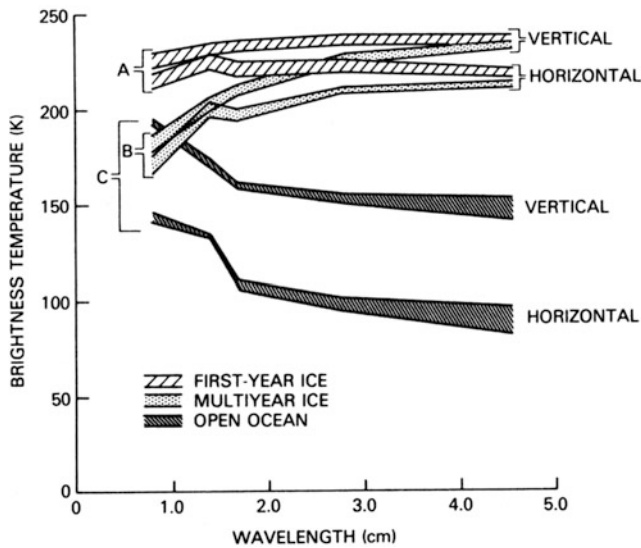
The use of satellite data from the visible and infrared Very High Resolution Radiometer (VHRR) began in 1966. In 1970, the National Oceanographic and Atmospheric Administration (NOAA) launched the first of a series of satellites with VHRR having an improved resolution of 1 km. In 1978, the first satellite carrying the improved Advanced Very High Resolution Radiometer (AVHRR) was launched. This series continues to this day.

Satellite reception at the Arctic and Antarctic Research Institute provides visible and infrared satellite images both from Russian (METEOR, OKEAN, RESURS) and US (NOAA) satellites. The OKEAN satellite also provides SLR and passive microwave data.

In December 1972, the National Aeronautics and Space Administration (NASA) launched the Electrically Scanning Microwave Radiometer (ESMR) on Nimbus 5. Until May 1977 this provided single channel horizontally polarized radiation at a frequency of 19 GHz. Its ability to operate in darkness and through cloud cover yielded the first comprehensive maps of polar sea ice extent for 1973–1976 (Zwally et al., 1983; Parkinson et al., 1987). The brightness temperature data gridded to 25 km (Parkinson et al., 1999) are available at: http://nsidc.org/data/docs/daac/nsidc0077_esmr_tbs.gd.html.

In October 1978, the Scanning Multichannel Microwave Radiometer (SMMR) was launched on Nimbus 7 and operated until August 1987. The instrument had three channels, two with dual polarization. Frequencies 18 and 37 GHz were used in various algorithms to derive sea ice concentrations for first-year and multiyear ice (Gloersen et al., 1992). The records continued with the Special Sensor Microwave Imager (SSM/I) on Defense Meteorological Program satellites. These instruments had five frequencies including 19 and 37 GHz.

Sea ice can be discriminated in the microwave regime through differences in the emissive characteristics between ice and ocean; in general, sea ice is more emissive than the open ocean. Use of combinations of frequencies



Cryosphere, Measurements and Applications,

Figure 1 Microwave brightness temperatures of sea ice and open water observed by Nimbus 7 SMMR in three regions of the Arctic, February 3–7, 1979. The regions A, B, and C are representative of first-year sea ice, multiyear ice, and open water, respectively. Hatched bands indicate $\pm$ one standard deviation about the mean (From Cavalieri et al., 1984; courtesy of the American Geophysical Union).

allows more accurate discrimination between ice and ocean as well as the ability to estimate fractional ice cover within regions of mixed ice and water. The NASA Team algorithm (Cavalieri et al., 1984) uses a polarization ratio and a gradient ratio.

The polarization ratio is:

$$PR_{[19V/H]} = \frac{T_B[19V] - T_B[19H]}{T_B[19V] + T_B[19H]}$$

The gradient ratio is:

$$GR_{[37V/19V]} = \frac{T_B[37V] - T_B[19V]}{T_B[37V] + T_B[19V]}$$

The PR is small for ice and large for water while the GR is small for first-year ice but large for multiyear ice. Figure 1 illustrates these points (Cavalieri et al., 1984). Combinations of PR and GR enable the T_B signatures to be interpreted as ice type and there are some eight or so algorithms in use for this purpose.

Another approach is used by the Bootstrap algorithm (Comiso, 1986), which employs linear combinations of 19 and 37 GHz frequencies at both horizontal and vertical polarizations to estimate fraction ice coverage. The NASA Team and Bootstrap, as well as other algorithms, require empirically derived “tie points,” or coefficients for pure surface types (100 % ice and 100 % water).

There are many uncertainties and limitations in using passive microwave data for sea ice detection. There are

errors due to ambiguous emissivity signals, particularly from surface melt water during summer and for thin ice. The available passive microwave frequencies can discriminate between at most three ice types, but often a region may have more than three unique microwave signatures. There can be erroneous ice retrievals over open water due to increased ocean surface emissivity from wind roughening. Atmospheric emission may be a factor in some conditions. Perhaps the major limitation is the low spatial resolution of passive microwave sensors, with footprints of 12–50 km. Thus, individual floes cannot be imaged and the ice edge location can be estimated to several kilometers accuracy at best. However, passive microwave data are a valuable source of sea ice information because it is sunlight independent and is generally not affected by clouds and other atmospheric sources. Also, passive microwave sensors have wide swaths and sun-synchronous orbits that provide frequent coverage of the polar regions. Thus, passive microwave data has provided a consistent and nearly complete daily record of sea ice conditions in both the Arctic and Antarctic since late 1978. Carsey (1992a) discusses the passive microwave properties of sea ice and the limitations of passive microwave remote sensing in more detail.

While the NASA Team and Bootstrap algorithm products are the most commonly used passive microwave sea ice estimates, several other algorithms have been developed, including the Cal/Val or AES York (Bjerkelund et al., 1990), Bristol (Hanna and Bamber, 2001), and Norsex (1983). All these algorithms make use of some combination of the 19 and 37 GHz channels. More recent algorithms have employed the higher frequency passive microwave channels on SSM/I and AMSR-E to obtain better spatial resolution and to resolve some of the surface ambiguities. These include the NASA Team 2 (Markus and Cavalieri, 2000) and the ARTIST (Spreen et al., 2008). The NASA Team 2 algorithm is used for the AMSR-E standard sea ice product (Comiso et al., 2003).

A comparison of the NASA Team and the Bootstrap algorithms is given by Comiso et al. (1997) and of the NASA Team, NASA Team 2, and the Bootstrap algorithms by Stroeve et al. (2001). Partington (2000) compared NASA Team and Cal/Val estimates and discussed the potential for a combined product using data fusion. Meier (2005) compares four different algorithms with AVHRR data. He finds that the smallest mean errors are from the Cal/Val and NASA Team-2 algorithms; the former tends to overestimate and the latter underestimate ice concentration relative to AVHRR in three near-marginal ice zones of the Arctic. Andersen et al. (2007) compared several algorithms and found significant differences in long-term trends depending on the algorithm, indicating potentially important influence of atmospheric and surface properties on the various sea ice algorithms.

Comiso and Nishio (2008) combine SSMR, SSM/I, and Advanced Microwave Scanning Radiometer (AMSR)-E data to show that the trend in ice area through 2006 in the Arctic is now slightly more negative at -4.0 ± 0.2 % per

decade, respectively, while the corresponding trend in the Antarctic remains slightly positive at $+1.7 \pm 0.3$ % per decade. The Arctic sea-ice cover, as revealed by microwave radiometry experienced a remarkable reduction in area in the summer of 2007, and again in 2012.

Carsey (1992a) provides a detailed and comprehensive treatment of sea ice remote sensing using both passive microwave and active radar remote sensing.

Passive microwave imagery is also useful for tracking large-scale sea ice motions using cross-correlated feature matching (e.g., Agnew et al., 1997; Kwok et al., 1998; Meier et al., 2000) or with a wavelet analysis approach (Liu and Cavalieri, 1998). SSM/I 85 GHz can estimate daily ice motions to within an RMS error of 6 km/day accuracy (Meier et al., 2000); averaging over longer time periods (Kwok et al., 1998) or using spatial interpolation (Meier et al., 2000) can reduce the error. Ice motion retrievals are limited during summer melt. Passive microwave sea ice motions have higher errors in the Antarctic (Kwok et al., 1998). The higher spatial resolution of AMSR-E can provide improved motion accuracy (Meier and Dai, 2006) and allow for better detection of summer motion through the use of the 18 GHz channel (Kwok, 2008). Because of the change in emissivity of sea ice during melt, passive microwave imagery is also useful for the determination of melt onset (Smith, 1998; Drobot and Anderson, 2001; Belchansky et al., 2004).

Radars that measure the power of the return pulse scattered back to the antenna can be used to derive geophysical parameters of the illuminated surface, or volume, based on the scattering principles of microwave electromagnetic radiation (Maurer, 2003). These instruments are known as “scatterometers.” The major instruments flown are the ESCAT, the European Space Agency’s (ESA) Earth Remote Sensing (ERS)-1 and -2 Active Microwave Instrument (C band, 3 GHz, V), the first of which operated between 1992 and 1996 and the second of which has been operating since 1996, and the NASA QuikSCAT SeaWinds instrument (Ku band, 13.6 GHz, V and H), flown from 1999 to present. Scatterometry is useful for measuring ice extent (Allen and Long, 2006; Anderson and Long, 2005; Remund and Long, 1999) and ice motion (Zhao et al., 2002; Haarpaintner, 2006). Resolution-enhanced scatterometer data using image reconstruction techniques have provided improved spatial resolution (Long et al., 1993), which potentially yields more precise ice edge position (Meier and Stroeve, 2008). Due to salinity differences between multiyear and first-year ice, scatterometer data can estimate seasonal and perennial ice coverage (e.g., Nghiem et al., 2007). Scatterometry is also useful for measuring the extent of snowmelt on sea ice due to its extreme sensitivity to the presence of liquid water.

Another active microwave sensor useful for sea ice studies is synthetic aperture radar. In contrast to scatterometers, it is an imaging radar that synthesizes images from multiple looks during the satellite’s motion in orbit to effectively create a large antenna and thus

obtain much higher spatial resolution. The Canadian RADARSAT-1 sensor has been providing SAR coverage of sea ice since 1995. RADARSAT-2 was launched in December 2007 and will take over SAR acquisition from RADARSAT-1. However, RADARSAT-2 is a purely commercial satellite and it is unlikely that data will be widely available to the science community. The resolution is high enough to capture small-scale ice motion and ice deformation events, allowing ice motion, ice age, ice volume, ice production, seasonal ice area to be estimated at fine spatial scales (Kwok and Cunningham, 2002; Kwok et al., 1995, 1998). The high-resolution, all-sky capabilities are particularly useful for operational analysis of sea ice, and SAR imagery is widely used by operational sea ice centers such as the Canadian Ice Service and the US National Ice Center (Bertoia et al., 2001). However, the narrow swath of SAR sensors limits repeat coverage to every 3–6 days in many regions of the Arctic. In addition, SAR imagery of sea ice can be difficult to interpret and efforts at automated analysis have been largely unsuccessful.

The determination of sea ice thickness remains a major challenge. Upward looking sonar (ULS) on submarines is the major source of ice thickness information in the Arctic (Rothrock et al., 1999). Ocean floor-mounted ULS have also provided ice thickness estimates but only in isolated locations (Vinje et al., 1998). Airborne electromagnetic induction instruments have been used experimentally and operationally since about 1990 (Haas and Eicken, 2001). Worby et al. (1999) describe its use over fast ice and pack ice in the Antarctic. More recently, autonomous underwater vehicles (AUVs) have been successfully employed to map the terrain of the underside of sea ice (Wadhams and Doble, 2008).

ERS radar altimetry has been used to estimate ice thickness in the Arctic (Laxon et al., 2003; Peacock and Laxon, 2004) and the Antarctic (Giles et al., 2008), but the coverage of the polar regions by the ERS orbit limits retrievals to the seasonal ice zones in the Arctic. Cryosat 2 with a radar altimeter was launched by ESA in April 2010 and is providing ice thickness and volume estimates for the Arctic (Laxon et al., 2012). The laser altimeter on the Ice, Cloud, and Land Elevation Satellite (ICESat) has been used to determine sea ice freeboard in the Arctic. Kwok et al. (2008) obtained average values of 43 cm over multiyear ice and 27 cm over first-year ice in February–March 2006. Uncertainties arise primarily through snow cover on the ice. Zwally et al. (2008) measured Antarctic ice thickness using ICESat.

There are considerable issues in determining sea ice thickness from altimeters and such measurements are still largely unvalidated. Because altimeters are measuring only freeboard, any errors in such estimates are magnified when deriving total ice thickness. The major uncertainty is snow thickness, which must be known to accurately account for its contribution to freeboard height (for laser altimeters that measure from the top of the snow surface) or for total freeboard density (for radar altimeters that

penetrate snow cover and measure from the snow/ice interface). However, while there is still considerable uncertainty in any absolute thickness estimates from altimeters, they appear to be able to capture seasonal and interannual variations well in sea ice thickness.

Lagrangian tracking of sea ice using passive microwave imagery and other sources provides a long history (since early 1980s) of ice age, a proxy for ice thickness (Maslanik et al., 2007); comparison with ICESat thickness indicates reasonable correlation between age and thickness on a basin scale up to ages of 9 years.

Ice sheets

Ice sheet remote sensing involves all wavelengths of the electromagnetic spectrum. Initial work was done with Landsat MSS images in the 1980s. Williams et al. (1982) assembled 4,270 Landsat scenes covering Antarctica to about 82 °S latitude. A Landsat MSS mosaic of the Ronne-Filchner ice shelf and Coats Land was georeferenced using measured ground control points by Sievers et al. (1989). A SAR mosaic from ERS-1 was georeferenced and is thought to be accurate to 50 m (Roth et al., 1993). The images are co-registered by either (1) matching fixed points such as nunataks projecting through the ice or (2) using the furnished coordinates based on orbital parameters. Note that georeferencing using the ephemeris data (furnished coordinates) has an advantage in the ice sheet interiors, where large areas are devoid of fixed rock outcrops. There are two methods to determine the glacial velocities: an interactive one in which crevasse patterns are visually traced (Lucchitta et al., 1993) and an autocorrelation program developed by Scambos et al. (1992).

Polar Pathfinder AVHRR products are available twice daily at 5 km resolution in Equal-Area Scalable Earth (EASE) Grid format (<http://nsidc.org/data/avhrr/>) for July 1981 through June 2005, and for shorter periods at 1.25 and 25 km resolution. The products include: clear-sky surface broadband albedo and skin temperature, solar zenith angle, surface type mask, cloud mask, orbit mask, and ice motion vectors (Maslanik et al., 1998; Scambos et al., 2000). Laine (2008) used these data to analyze albedo changes in the Antarctic. All sectors show slight increasing spring–summer albedo trends. The steepest ice sheet albedo trend of $0.0019 \pm 0.0009/\text{year}$ is found in the Ross Sea sector. The steepest sea ice albedo trend of $0.0044 \pm 0.0017/\text{year}$ occurs in the Pacific Ocean sector.

Multi-angle Imaging SpectroRadiometer (MISR) data have been used to determine the ice albedo over Greenland (Nolin et al., 2001). Stroeve and Nolin (2002) use two different methods to derive the snow albedo: one based on the spectral information and one utilizing the angular information from the MISR instrument. The latter method is based on a statistical relationship between in situ albedo measurements and the MISR red channel reflectance at all MISR viewing angles and is found to

give good agreement with the ground-based measurements. The spectral method is more sensitive to instrument calibration and shows bidirectional reflectance distribution function models and narrowband-to-broadband albedo relationships. The MISR sensor's ability to map glacier facies and roughness was explored and documented by Nolin et al. (2002).

The VELMAP project at NSIDC (<http://nsidc.org/data/velmap/>) aims to compile all ice flow data for the Antarctic continent. The project includes Landsat 7 and ASTER ice velocity maps while the second and third Radarsat Antarctic Mapping Missions provide data north of 80 °S. There are currently more than 130,000 ice vectors in the database.

Changes in the West Antarctic ice sheet since 1963 have been identified using formerly classified Corona images and recent data by Bindschadler and Vornberger (1998). They find that ice stream B has widened at a rate much faster than expected and its movement also slowed down. The available data are reviewed and described in Bindschadler and Seider (1998). A mosaic of the 1962–1963 Antarctic coast and ice shelves was compiled, and a grounding line for that time was calculated, from these data (Kim et al., 2006).

In 1997, the Canadian RADARSAT-1 satellite was rotated in orbit, so that its synthetic aperture radar (SAR) antenna looked south toward Antarctica. This permitted the first high-resolution mapping of the entire Antarctic continent to be accomplished in less than 3 weeks. Swath images representing calibrated radar backscatter data have been assembled into 90 tiles at 25 m resolution and an image mosaic available at 125 m to 1 km resolutions. The RADARSAT Image Map of Antarctica, 1999, is the product of the RADARSAT-1 Antarctic Mapping Project (RAMP). The mosaic provides a detailed look at ice sheet morphology, rock outcrops, research infrastructure, the coastline, and other features. Accompanying this mosaic is the high-resolution RAMP Digital Elevation Model (DEM) that combines topographic data from a variety of non-SAR sources to provide consistent coverage of all of Antarctica. Version 2 improves upon the original version by incorporating new topographic data, error corrections, extended coverage, and other modifications. The RAMP DEM is gridded at 200, 400, and 1,000 m.

There are large-scale spatial variations in radar brightness. The bright portion of Marie Byrd Land and the eastern Ross Ice Shelf probably represents the region where significant melting and refreezing occurred during an early 1990s melt event. Most of the coastal areas and much of the Antarctic Peninsula appear bright also because of summer melt. Remaining strong variations in radar brightness are poorly understood. Thousands of kilometer long curvilinear features across East Antarctica appear to follow ice divides separating the large catchment areas. On an intermediate scale, the East Antarctic Ice Sheet appears to be very “rough.” The texturing is probably due to the flow of the ice over the Wilkes subglacial basin located in George V Land. There the imagery shows subtle

rounded shapes similar in appearance to the signature of subglacial lakes such as Lake Vostok. In Queen Maud Land, East Antarctica, there are extensive ice stream and ice stream-like features. An enormous ice stream fed by a funnel-shaped catchment, reaching at least 800 km into East Antarctica, feeds Recovery Glacier, which enters the Filchner Ice Shelf. Megadunes and wind glaze regions were first identified as local features in Landsat TM, but are fully visible for the first time in the RAMP mosaic. Megadunes cover approximately one million square kilometers of the plateau surface (Fahnestock et al., 2006).

Using interferometry with SAR imagery obtained during the 1997 Antarctic Mapping Mission, ice velocity vectors were obtained over the East Antarctic ice streams. The upstream velocity of the Recovery Glacier is about 100 m/year but near the grounding line there is a local peak velocity of about 900 m/year.

Two digital image maps of surface morphology and optical snow grain size that cover the Antarctic continent and its surrounding islands have been prepared by NSIDC and the University of New Hampshire (Haran et al., 2005). The MODIS Mosaic of Antarctica (MOA) image maps are derived from composites of 260 MODIS orbit swaths. The MOA provides a cloud-free view of the ice sheet, ice shelves, and land surfaces, and a quantitative measure of optical snow grain size for snow- or ice-covered areas. Scambos et al. (2007) use MOA to provide continent-wide surface morphology and snow grain size.

Recently, the USGS with the British Antarctic Survey, NASA, and the National Science Foundation have produced a Landsat Image Mosaic of Antarctica (LIMA) from over 1,000 scenes of ETM+ to latitude 82.5°S at 15 m resolution (http://lima.usgs.gov/view_lima.php). Both the mosaic and the individual original Landsat images used in the compilation (which are nearly completely cloud-free scenes of the surface) are available online for free.

For Greenland, there has been albedo mapping with AVHRR (Stroeve et al., 1997) and MODIS (Liang et al., 2005). Ice velocity in ice streams has been derived from interferometric SAR (InSAR) (Joughin et al., 2000) and C-band SAR data from ERS-1 have been used to map the different snow and ice facies of the ice sheet (Fahnestock et al., 1993). Snow melt over the ice sheet has been mapped using SSM/I data by Abdalati and Steffen (1996) and using SMMR and SSM/I data by Mote and Anderson (1995). The latter used a threshold value of the 37 GHz brightness temperature while Abdalati and Steffen used a cross-polarized gradient ratio (XPGR), which is a normalized difference between the 19 GHz horizontally polarized and 37 GHz vertically polarized brightness temperatures. The threshold of $XPGR = -0.025$ is used to classify dry versus wet snow.

Changes in mass balance of the two major ice sheets have recently been derived from a variety of satellite measurements including the Gravity Recovery and Climate Experiment (GRACE) (Rignot and Thomas, 2002).

They show that the Greenland ice sheet is losing mass by near-coastal thinning. The West Antarctic ice sheet, with thickening in the west and thinning in the north, is probably thinning overall. The mass imbalance of the East Antarctic ice sheet is likely to be small, but even its sign is uncertain. The main objective of the laser altimetry data obtained from the Geoscience Laser Altimeter System (GLAS) on the Ice, Cloud, and land Elevation Satellite (ICESat) was to measure ice sheet elevations and changes in elevation (Schutz et al., 2005). GLAS provided global coverage between 86°N and 86°S. ICESat has seen wide use as a tool in investigating specific processes or regions, via profile comparisons over time. Among several results are the discovery of rapidly changing subglacial lake systems, which appear to fill and drain on short (month to year) timescales (Fricker et al., 2007; Shepherd and Wingham, 2007). Using satellite-derived surface elevation and velocity data, Howat et al. (2007) found large short-term variations in recent ice discharge and mass loss at two of Greenland's largest outlet glaciers (Howat et al., 2007).

Ice shelves

Iceberg calving from the Ross and Filchner ice shelves has been observed via satellite by the National Ice Center (NIC) originally using DMSP Optical Line Scanner (OLS) with 2.7 km resolution and AVHRR and subsequently MODIS and Envisat. Ballantyne and Long (2002) use scatterometer data to show the increasing numbers of icebergs detected since 1976. Antarctic icebergs are designated by the NIC as originating in one of four quadrants: A 0–90°W, B 90–180°W, C 180–90°E, and D 90–0°E. Iceberg B-15 broke off the Ross Ice Shelf in late March, 2000. Among the largest ever observed, this iceberg was approximately 270 km long $\times$ 40 km wide, nearly as large as the state of Connecticut.

MODIS satellite imagery revealed that the northern section of the 220 m-thick Larsen B ice shelf, on the eastern side of the Antarctic Peninsula, shattered and separated from the continent. A total of about 3,250 km² of shelf area disintegrated in a 35 day period beginning on January 31, 2002. The shattered ice formed a plume of thousands of icebergs adrift in the Weddell Sea. Over the last 5 years, the shelf has lost a total of 5,700 km², and is now about 40 % the size of its previous minimum stable extent. Glasser and Scambos (2008) found that ice-shelf breakup is not controlled simply by climate. A number of other atmospheric, oceanic, and glaciological factors are involved. The location and spacing of fractures on the ice shelf such as crevasses and rifts are very important because they determine the strength of the ice shelf.

On Ellesmere Island, Arctic Canada, the Ward Hunt Ice Shelf has undergone similar rapid changes since 2000. SAR imagery revealed an extensive serpentine crack, and secondary fractures, in 2002 (Mueller et al., 2003). In summer 2005, the Ayles Ice Shelf broke off forming an ice island (Copeland et al., 2007).

Icebergs

Icebergs were originally tracked by airborne reconnaissance in the Canadian Eastern Arctic and sub-Arctic. During 1960–1968, the International Ice Patrol used visual aerial reconnaissance. From 1963 to 1982, the Ice Patrol made iceberg survey flights north along the Labrador coast and into Baffin Bay. Side-looking airborne radar (SLAR) was introduced in 1983 and continues to be used. Following the advent of Radarsat-1 SAR (Power et al., 2001) in November 1995, the costly airborne surveys were discontinued by the Canadian Ice Service.

The Antarctic Meteorological Research Center at the University of Wisconsin-Madison provides near-real-time and archived imagery of Antarctic icebergs (<http://amrc.ssec.wisc.edu/index.html>). The images are from NOAA Polar Orbiting visible and IR band data. Brigham Young University, Provo, UT, produced enhanced resolution scatterometer backscatter images during July–September 1978 (from Seasat), July 1996–June 1997 (from NSCAT), 1992–2001 (from ERS-1 and 2), and June 1999 up to present (from QuikSCAT). Images were obtained from the Scatterometer Climate Record Pathfinder (SCP) project (Stuart et al., 2007). The initial position for each iceberg is located based on either (1) a position reported by the National Ice Center's web page (<http://www.natice.noaa.gov/>) or (2) by the sighting of a moving iceberg in a time series of scatterometer images. Ballantyne and Long (2002) used the archive to produce a long-term analysis of Antarctic iceberg activity.

Frozen ground

Frozen ground may be seasonal or perennial (permafrost). Permafrost is beneath the surface and not readily amenable to direct remote sensing (Zhang et al., 2004). However, the near-surface soil freeze/thaw status can be determined using satellite remote sensing data. SAR provides information on the timing, duration, and spatial progression of near-surface freeze/thaw in autumn and spring, for example. Freezing results in a large increase in the dielectric of soil and vegetation, which causes a large decrease in L-band (15–30 cm wavelength) and C-band (3.75–7.5 cm wavelength) radar backscatter (3 dB). Way et al. (1997) used the ERS-1 C-band instrument over central Canada, for example, to detect this. Passive microwave radiation (PMR) data offer similar information at lower spatial resolution. Frozen soils relative to unfrozen soils exhibit (1) lower thermal temperatures, (2) higher emissivity, and (3) lower brightness temperatures. The PMR algorithm for frozen soils is:

$$\frac{\partial}{\partial f} T_B(f) \leq P_{SG}$$

and

$$T_{B(37V)} \leq P_D$$

where the spectral gradient is in $K \text{ GHz}^{-1}$ and $T_{B(37V)}$ is in K. P_{SG} and P_D are the cutoff spectral gradient and brightness temperature, respectively. Based on these equations, surfaces can be classified as frozen, dry (and hot), wet (and cool), and mixed (Zuerndorfer and England, 1992). A frozen surface has low brightness temperature (37 GHz) and a relatively low negative spectral gradient. Zhang and Armstrong (2001) and Zhang et al. (2003) analyzed soil freeze/thaw status over the contiguous United States and southern Canada in winter 1997/1998. They used a negative spectral gradient and a threshold value of $P_{37} = 258.2 \text{ K}$. They found that almost 80 % of the time, the near-surface soil was frozen before snow accumulated on the ground. They applied the validated frozen soil algorithm to investigate near-surface soil freeze/thaw status from 1978 through 2003 over the Northern Hemisphere (Zhang and Armstrong, 2003). The long-term average maximum area extent of seasonally frozen ground, including the active layer over permafrost, is approximately 50.5 % of the landmass in the Northern Hemisphere. Preliminary results indicate that the extent of seasonally frozen ground has decreased about 15–20 % during the past few decades.

Smith et al. (2004) developed a freeze/thaw algorithm for SSM/I data and applied it for high northern latitudes for 1988–2002. They found a trend toward later autumn freeze-up in evergreen conifer forests in North America by 3.1 ± 1.2 days/decade while in Eurasia there was a trend toward earlier thaw dates in tundra (-3.3 ± 1.8 days/decade) and larch biomes (-4.5 ± 1.8 days/decade). Despite the trend toward earlier thaw dates in Eurasian larch forests, the growing season did not increase in length because of parallel changes in timing of the fall freeze (-5.4 ± 2.1 days/decade).

Conclusions

Remote sensing of the cryosphere has made major advances over the last three decades and has become an indispensable tool for cryospheric research, given the remote locations and hostile environments involved in the spatial distribution of many cryospheric variables. Beginning with aerial photography of glaciers, optical remote sensing then came into use in the 1960s–1970s for mapping snow cover and sea ice. The advent of passive microwave remote sensing in the 1970s eliminated most of the problems of cloud cover and illumination, and this advantage continued with the much higher resolution airborne and satellite radar data in the 1980s. Spatial resolution of optical sensors became much enhanced in the 1990s, and the 2000s have seen the use of satellite laser altimeter and gravity measurements. Techniques have also advanced with the direct production of DEMs from satellite data and ice motion studies using Interferometric SAR.

Bibliography

- Abdalati, W., and Steffen, K., 1996. Passive microwave-derived snow melt regions on the Greenland Ice Sheet. *Geophysical Research Letters*, **22**, 787–790.
- Agnew, T. A., Le, H., and Hirose, T., 1997. Estimation of large-scale sea-ice motion from SSM/I 85.5 GHz imagery. *Annals of Glaciology*, **25**, 305–311.
- Aizen, V. B., Aizen, E. M., and Kuzmichenok, V. A., 2007. Glaciers and hydrological changes in the Tien Shan: simulation and prediction. *Environmental Research Letters*, **2**, 045019, 10 pp.
- Allen, J. R., and Long, D. G., 2006. Microwave observations of daily Antarctic sea-ice edge expansion and contraction rates. *IEEE Geoscience and Remote Sensing Letters*, **3**(1), 54–58.
- Altmaler, A., and Kany, C., 2002. Digital surface model generation from CORONA satellite images. *Journal of Photogrammetry and Remote Sensing*, **56**, 221–235.
- Andersen, S., Tonboe, R., Kaleschke, L., Heygster, G., and Pedersen, L. T., 2007. Intercomparison of passive microwave sea ice concentration retrievals over high-concentration Arctic sea ice. *Journal of Geophysical Research*, **112**, C08004, doi:10.1029/2006JC003543.
- Anderson, H. S., and Long, D. G., 2005. Sea ice mapping method for SeaWinds. *IEEE Transactions on Geoscience and Remote Sensing*, **43**(3), 647–657.
- Arctic and Antarctic Research Institute, 2007. Sea ice charts of the Russian arctic in gridded format, 1933–2006. In Smolyanitsky, V., Borodachev, V., Mahoney, A., Fetterer, F., and Barry, R. (eds.), *Digital Media*. Boulder, CO: National Snow and Ice Data Center.
- Armstrong, R. L., et al., 2005a. *Global Monthly EASE-Grid Snow Water Equivalent Climatology*. Digital Media. Boulder, CO: National Snow and Ice Data Center.
- Armstrong, R., et al., 2005b. *GLIMS Glacier Database*. Digital Media. Boulder, CO: National Snow and Ice Data Center.
- Asmus, V., 2003. Russian environmental satellites: current status and development perspectives. In *Committee on Earth Observation Satellites, 17th Plenary Meeting Colorado Springs*, November 19–20, 2003, Colorado Springs, CO, Item 19.5.
- Ballantyne, J., and Long, D. G., 2002. A multidecadal study of the number of Antarctic icebergs using scatterometry data. In *Geoscience and Remote Sensing Symposium*, 2002. *IGARSS'02*, Vol. 5, pp. 3029–3031.
- Barry, R. G., and Gan, T. Y., 2011. *The global cryosphere: past, present and future*. Cambridge: Cambridge University Press, 498 pp.
- Bayr, K. J., Hall, D. K., and Kovalick, W. M., 1994. Observations on glaciers in the eastern Austrian Alps using satellite data. *International Journal of Remote Sensing*, **15**, 1733–1742.
- Belchansky, G. I., Douglas, D. C., and Platonov, N. G., 2004. Duration of the Arctic melt season: regional and interannual variability, 1979–2001. *Journal of Climate*, **17**, 67–80.
- Bertoia, C., Manore, M., and Andersen, H. S., 2001. Mapping ice covered waters from space. *Backscatter*, **12**, 14–21.
- Bindschadler, R., and Seider, W., 1998. *Declassified intelligence satellite photography (DISP) coverage of Antarctica*. NASA Technical Memorandum 1998–206879, 38 pp.
- Bindschadler, R., and Vornberger, P., 1998. Changes in the West Antarctic Ice Sheet since 1963 from declassified satellite photography. *Science*, **279**, 689–692.
- Bishop, M. P., et al., 2004. Global Land Ice Measurements from Space (GLIMS): Remote sensing and GIS investigations of the Earth's cryosphere. *Geocarto International*, **19**(2), 57–84.
- Bjerkelund, C. A., Ramseier, R. O., and Rubinstein, I. G., 1990. Validation of the SSM/I and AES/York algorithms for sea ice parameters. In Ackley, S. F., and Weeks, W. F. (eds.), *Sea Ice Properties and Processes*. Hannover, NH: USACE. U.S. Army Cold Regions Research and Engineering Laboratory Monograph, Vol. 90–91, pp. 206–208.
- Borodachev, B. E., and Shilnikov, V. I., 2003. *Istoriya L'dovoi Aviatsonnoi Razedki v Arktikei na Zamerzayushchikh Moryakh Rossii (1924–1993) [The History of Aerial Ice Reconnaissance in the Arctic and Ice-covered Seas of Russia, 1924–1993]*. St. Petersburg: Gidrometeoizdat, 441 pp.
- Carsey, F. (ed.), 1992a. *Microwave Remote Sensing of Sea Ice*. Washington, DC: American Geophysical Union. Geophysical Monograph, Vol. 68, 478 pp.
- Carsey, F., 1992b. Remote sensing of ice and snow: review and status. *International Journal of Remote Sensing*, **13**, 5–11.
- Cavalieri, D. J., Gloersen, P., and Campbell, W. J., 1984. Determination of sea ice parameters with Nimbus 7 SMMR. *Journal of Geophysical Research*, **89**(C3), 5355–5369.
- Chang, A. T. C., Foster, J. L., and Hall, D. K., 1987. Nimbus-7 SMMR derived global snow cover parameters. *Annals of Glaciology*, **9**, 39–44.
- Chang, A. T. C., Foster, J. L., and Hall, D. K., 1990. Satellite sensor estimates of northern hemisphere snow volume. *International Journal of Remote Sensing*, **11**, 167–171.
- Chang, A. T. C., et al., 2003. Global SWE monitoring using AMSR-E data. In *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, 2003, pp. 680–682.
- Cline, D., et al., 1999. *Cold Land Processes Mission (EX-7) Science and Technology Implementation Plan*. Report on the NASA Post-2002 Land Surface Hydrology Planning Workshop, Irvine, CA.
- Comiso, J. C., 1986. Characteristics of Arctic winter sea ice from satellite multispectral microwave observation. *Journal of Geophysical Research*, **91**(C1), 975–994.
- Comiso, J. C., and Nishio, F., 2008. Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data. *Journal of Geophysical Research*, **113**(C2), C02S07.
- Comiso, J. C., Cavalieri, D., Parkinson, C., and Gloersen, P., 1997. Passive microwave algorithms for sea ice concentrations: a comparison of two techniques. *Remote Sensing of Environment*, **60**(3), 357–384.
- Comiso, J. C., Cavalieri, D. J., and Markus, T., 2003. Sea ice concentration, ice temperature, and snow depth using AMSR-E data. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 243–252.
- Copeland, L., Mueller, D. R., and Weir, L., 2007. Rapid loss of the Ayles Ice Shelf, Ellesmere Island, Canada. *Geophysical Research Letters*, **34**, L21501.
- Csatho, B., et al., 2000. Investigating the longterm behavior of Greenland outlet glaciers using high-resolution satellite imagery. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARRS)*, 2002, pp. 1047–1150.
- Dedieu, J. F., and et al., 2003. Glacier mass balance determination by remote sensing in the French Alps: Progress and limitation for time series monitoring. In *Proceedings of the Fourth International Geoscience and Remote Sensing Symposium (IGARSS'03)*, pp. 2602–2604.
- Dozier, J., and Marks, D., 1987. Snow mapping and classification from Landsat Thematic Mapper data. *Annals of Glaciology*, **9**, 97–103.
- Dozier, J., and Painter, T. H., 2004. Multispectral and hyperspectral remote sensing of alpine snow properties. *Annual Reviews of Earth and Planetary Sciences*, **32**, 465–494.
- Drobot, S. D., and Anderson, M. R., 2001. An improved method for determining snowmelt onset dates over Arctic sea ice using scanning multichannel microwave radiometer and special sensor microwave/imager data. *Journal of Geophysical Research*, **106** (D20), 24,033–24,050.
- Duguay, C. R., and Lafleur, P. M., 2003. Estimating depth and ice thickness of shallow subarctic lakes using spaceborne optical

- and SAR data. *International Journal of Remote Sensing*, **24**(3), 475–489.
- Dunbar, M. R., and Greenway, K. R., 1956. *Arctic Canada from the Air*. Ottawa: Defense Research Board, 541 pp.
- Fahnestock, M., et al., 1993. Greenland Ice Sheet surface properties and ice dynamics from ERS-1 SAR imagery. *Science*, **262**(5139), 1530–1534.
- Fahnestock, M., Scambos, T., Haran, T., and Bauer, R., 2006. *AWS data: characteristics of snow megadunes and their potential effect on ice core interpretation*. Digital media. Boulder, CO: National Snow and Ice Data Center.
- Foster, J. L., et al., 1980. Snowpack monitoring in North America and Eurasia using passive microwave satellite data. *Remote Sensing of Environment*, **10**, 285–298.
- Foster, J., Chang, A., and Hall, D., 1997. Comparison of snow mass estimates from a prototype passive microwave snow algorithm, a revised algorithm and a snow depth climatology. *Remote Sensing of Environment*, **62**, 132–142.
- Fricker, H. A., et al., 2007. An active subglacial water system in West Antarctica mapped from space. *Science*, **315**(5818), 1544–1548.
- Giles, K. A., Laxon, S. W., and Worby, A. P., 2008. Antarctic sea ice elevation from satellite radar altimetry. *Geophysical Research Letters*, **35**, L03503, doi:10.1029/2007GL031572.
- Glasser, N. F., and Scambos, T. A., 2008. A structural glaciological analysis of the 2002 Larsen B ice shelf collapse. *Journal of Glaciology*, **54**(184), 3–16.
- Gloersen, P., Campbell, W. J., Cavalieri, D. J., Comiso, J. C., Parkinson, C. L., and Zwally, H. J., 1992. *Arctic and Antarctic Sea Ice, 1978–1987: Satellite Passive-Microwave Observations and Analysis*. Washington, DC: NASA.
- Goodison, B. E., and Walker, A. E., 1994. Canadian development and use of snow cover information from passive microwave satellite data. In Choudhury, B. (ed.), *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*. Utrecht: VSP VB, pp. 245–262.
- Haarpaintner, J., 2006. Arctic-wide operational sea ice drift from enhanced resolution QuikScat/SeaWinds scatterometry and its validation. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 1433–1443.
- Haas, C., and Eicken, H., 2001. Interannual variability of summer sea ice thickness in the Siberian and central Arctic under different atmospheric circulation regimes. *Journal of Geophysical Research*, **106**(C3), 4449–4462.
- Hall, D. K., and Riggs, G. A., 2007. Accuracy assessment of the MODIS snow products. *Hydrological Processes*, **21**, 1534–1547.
- Hall, D. K., Williams, R. S., Jr., and Bayr, K. J., 1992. Glacier recession in Iceland and Austria. *Eos, Transactions of the American Geophysical Union*, **73**, 129, 135 and 141.
- Hanna, E., and Bamber, J., 2001. Derivation and optimization of a new Antarctic sea-ice record. *International Journal of Remote Sensing*, **22**, 113–139.
- Haran, T. (ed.), 2003. *CLPX-Satellite: MODIS radiances, reflectances, snow cover and related grids*. Digital Media. Boulder, CO: National Snow and Ice Data Center.
- Haran, T., et al., 2005 updated 2006. *MODIS mosaic of Antarctica (MOA) image map*. Digital media. Boulder, CO: National Snow and Ice Data Center.
- Helfrich, S., et al., 2007. Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice Mapping System (IMS). *Hydrological Processes*, **21**, 1576–1586.
- Howat, I. M., Joughin, I., and Scambos, T. A., 2007. Rapid changes in ice discharge from Greenland outlet glaciers. *Science*, **315**(5818), 1559–1561.
- Johannessen, O. M., et al., 2007. *Remote Sensing of Sea Ice in the Northern SeaRoute: Studies and Applications*. Chichester: Springer, Praxis, p. 472.
- Joughin, I. R., Fahnestock, M. A., and Bamber, J. L., 2000. Ice flow in the northeast Greenland ice stream. *Annals of Glaciology*, **31**, 141–146.
- Khromova, T. E., et al., 2006. Changes in glacier extent in the eastern Pamir, Central Asia, determined from historical data and ASTER imagery. *Remote Sensing of Environment*, **102**, 24–32.
- Kim, K., Jezek, K., and Liu, H., 2006. Orthorectified image mosaic of Antarctica from 1963 Argon satellite photography: image processing and glaciological applications. *International Journal of Remote Sensing*, **28**(23), 5357–5373.
- Kukthuroja, N., White, M. A., and Luecke, C., 2006. *Using Daily MODIS 250 m Data to Monitor Lake Ice Out*. Department of Watershed Sciences, Utah State University, 5 pp.
- Kwok, R., 2008. Summer sea ice motion from the 18 GHz channel of AMSR-E and the exchange of sea ice between the Pacific and Atlantic sectors. *Geophysical Research Letters*, **35**, L03504, doi:10.1029/2007GL032692.
- Kwok, R., and Cunningham, G. F., 2002. Seasonal sea ice area and volume production of the Arctic Ocean: November 1996 through April 1997. *Journal of Geophysical Research*, **107**, 8038, doi:10.1029/2000JC000469.
- Kwok, R., et al., 1995. Determination of the age distribution of sea ice from Lagrangian observations of ice motion. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 392–400.
- Kwok, R., et al., 1998. Sea ice motion from satellite passive microwave imagery assessed with ERS SAR and buoy motions. *Journal of Geophysical Research*, **103**(C4), 8191–8214.
- Kwok, R., et al., 2008. Ice, Cloud, and Land Elevation Satellite (ICESat) over Arctic sea ice: retrieval of freeboard. *Journal of Geophysical Research*, **112**(1–19), C12013.
- Laine, V., 2008. Antarctic ice sheet and sea ice regional albedo and temperature change, 1981–2000, from AVHRR Polar Pathfinder data. *Remote Sensing of Environment*, **112**(3), 646–667.
- Lauriol, B., et al., 1986. The residual snow cover in the Canadian Arctic in July: a means to evaluate the regional maximum snow depth in winter. *Arctic*, **39**, 309–315.
- Laxon, S. W., Peacock, N., and Smith, D., 2003. High interannual variability of sea ice thickness in the Arctic region. *Nature*, **425**, 947–950.
- Laxon, S. W., et al., 2012. CryoSat 2 estimates of Arctic ice thickness and volume. *Geophysical Research Letters*, **40**, 732–737.
- Leconte, R., and Klassen, P. D., 1991. Lake and river ice investigations in northern Manitoba using airborne SAR imagery. *Arctic*, **44**(Suppl. 1), 153–163.
- Leshkevich, G. A., Ngien, S. V., and Kwok, R. 1998. Algorithm development for satellite synthetic aperture radar (SAR) classification and mapping of Great Lakes ice cover. In *Proceedings, IEEE International Geoscience and Remote Sensing Symposium (IGARSS'98)*, Seattle, WA, 3 pp.
- Liang, S., Stroeve, J., and Box, J., 2005. MODIS-derived Greenland ice sheet albedo: the improved direct estimation algorithm. *Journal of Geophysical Research*, **110**, doi:10.1029/2004JD005493.
- Liu, A. K., and Cavalieri, D. J., 1998. On sea ice drift from the wavelet analysis of the Defense Meteorological Satellite Program (DMSP) Special Sensor Microwave Imager (SSM/I) data. *International Journal of Remote Sensing*, **19**, 1415–1423.
- Long, D. G., Hardin, P. J., and Whiting, P. T., 1993. Resolution enhancement of spaceborne scatterometer data. *IEEE Transactions on Geoscience and Remote Sensing*, **31**, 700–715.
- Lucchitta, B. K., et al., 1993. Antarctic glacial tongue velocities from Landsat images: first results. *Annals of Glaciology*, **17**, 356–366.
- Mahoney, A. R., et al., 2008. Observed sea ice extent in the Russian Arctic, 1933–2006. *Journal of Geophysical Research*, **113**, C110015.

- Markus, T., and Cavalieri, D., 2000. An enhancement of the NASA team sea ice algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 1387–1398.
- Martinec, J., and Rango, A., 1987. Interpretation and utilization of areal snow cover data from satellites. *Annals of Glaciology*, **9**, 166–169.
- Maslanik, J., et al., 1998. AVHRR-based polar pathfinder products for modeling applications. *Annals of Glaciology*, **25**, 388–392.
- Maslanik, J. A., et al., 2007. A younger, thinner Arctic ice cover: increased potential for rapid, extensive sea-ice loss. *Geophysical Research Letters*, **34**, L24501, doi:10.1029/2007GL032043.
- Maurer, J., 2003. Spaceborne scatterometry of the cryosphere: a review. http://cires.colorado.edu/maurerj/scatterometry/scatterometry_cryosphere
- McDonald, R. M., 1995. Opening the cold war sky to the public: declassifying satellite reconnaissance imagery. *Photogrammetric Engineering and Remote Sensing*, **61**, 485–490.
- Meier, W. M., 2005. Comparison of passive microwave ice concentration algorithm retrievals with AVHRR imagery in Arctic peripheral seas. *IEEE Transactions in Geoscience and Remote Sensing*, **43**(6), 1324–1337.
- Meier, W. N., and Dai, M., 2006. High-resolution sea-ice motions from AMSR-E imagery. *Annals of Glaciology*, **44**, 353–356.
- Meier, W. N., and Stroeve, J., 2008. Comparison of sea-ice extent and ice-edge location estimates from passive microwave and enhanced-resolution scatterometer data. *Annals of Glaciology*, **48**, 65–70.
- Meier, W. N., Maslanik, J. A., and Fowler, C. W., 2000. Error analysis and assimilation of remotely sensed ice motion within an Arctic sea ice model. *Journal of Geophysical Research*, **105**, 3339–3356.
- Mote, T. L., and Anderson, M. R., 1995. Variations in snowpack melt on the Greenland ice sheet based on passive-microwave measurements. *Journal of Glaciology*, **41**(137), 51–60.
- Mueller, D. R., Vincent, W. F., and Jeffries, M. O., 2003. Ice shelf break-up and ecosystem loss in the Canadian High Arctic. *Eos, Transactions American Geophysical Union*, **84**(49), 548–552.
- Nghiem, S. V., et al., 2007. Rapid reduction of Arctic perennial sea ice. *Geophysical Research Letters*, **34**, L19504, doi:10.1029/2007GL031138.
- Nolin, A. W., et al., 2001. Cryospheric applications of MISR data. In *Proceedings of the Third IEEE International Geoscience and Remote Sensing Symposium (IGARRS) 2001*. New York: IEEE, pp. 1219–1221.
- Nolin, A., Fetterer, F., and Scambos, T., 2002. Surface roughness characterizations of sea ice and ice sheets: case studies with MISR data. *IEEE Transactions of Geophysics and Remote Sensing*, **40**(7), 1605–1615.
- NORSEX Group, 1983. Norwegian remote sensing experiment in a marginal ice zone. *Science*, **220**(4599), 781–787.
- Painter, T. H., Roberts, D. A., Green, R. O., and Dozier, J., 1998. The effect of grain size on spectral mixture analysis of snow-covered area with AVIRIS data. *Remote Sensing of Environment*, **65**, 320–332.
- Painter, T. H., Dozier, J., Roberts, D. A., Davis, R. E., and Green, R. O., 2003. Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data. *Remote Sensing of Environment*, **85**, 64–77.
- Parkinson, C. L., et al., 1987. *Antarctic sea ice, 1973–1976: satellite passive-microwave observations*. Washington, DC: NASA, SP 489, 296 pp.
- Parkinson, C., Comiso, J., and Zwally, H. J., 1999. *Nimbus-5 ESMR daily polar gridded brightness temperatures*. Digital media. Boulder, CO: National Snow and Ice Data Center.
- Partington, K. C., 2000. A data fusion algorithm for mapping sea-ice concentrations from Special Sensor Microwave/Imager data. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 1947–1958.
- Paul, F., 2000. Evaluation of different methods for glacier mapping using Landsat TM. In *Proceedings, EARSeL-SIG Workshop*. Dresden: Land ice and Snow.
- Peacock, N. R., and Laxon, S. W., 2004. Sea surface height determination in the arctic ocean from ERS altimetry. *Journal of Geophysical Research*, **109**, C07001, doi:10.1029/2001JC001026.
- Power, D., et al., 2001. Iceberg detection capabilities of RADARSAT synthetic aperture radar. *Canadian Journal of Remote Sensing*, **27**(5), 476–486.
- Racoviteanu, A., et al., 2008. Decadal changes in glacier parameters in the Cordillera Blanca, Peru, derived from remote sensing. *Journal of Glaciology*, **56**(186), 499–510.
- Ramsay, B. H., 1998. Interactive Multisensor Snow and Ice Mapping System (IMS). *Hydrological Processes*, **12**, 1537–1546.
- Raup, B., et al., 2007. The GLIMS geospatial glacier database: a new tool for studying glacier change. *Global and Planetary Change*, **56**, 101–110.
- Remund, Q. P., and Long, D. G., 1999. Sea ice extent mapping using Ku band scatterometer data. *Journal of Geophysical Research*, **104**(C5), 11,515–11,527.
- Rignot, E., and Thomas, R. H., 2002. Mass balance of polar ice sheets. *Science*, **297**(5586), 1502–1506.
- Roth, A., et al., 1993. Experiences with ERS-1 SAR compositional accuracy. *Proceedings of the Third IEEE Transactions Geoscience. Remote Sensing, IGARRS Symposium*, 1993, Tokyo, Japan, pp. 1450–1452.
- Rothrock, D. A., Yu, Y., and Maykut, G. A., 1999. Thinning of the Arctic sea-ice cover. *Geophysical Research Letters*, **26**(23), 3469–3472.
- Rott, H., and Markl, G., 1989. Improved snow and glacier monitoring by the Landsat Thematic Mapper. In *Proceedings of a Workshop on Landsat Thematic Mapper Applications*, ESA-SP-1102, European Space Agency, pp. 3–12.
- Rott, H., and Nagler, T., 1993. Capabilities of ERS-1 SAR for snow and glacier monitoring in alpine areas. In *Proceedings, Second ERS-1 Symposium*. Noordwijk: ESA Publications Division, ESA SP-361, pp. 1–6.
- Scambos, T. A., et al., 1992. Application of image cross-correlation to the measurement of glacial velocity using satellite image data. *Remote Sensing of Environment*, **42**, 177–186.
- Scambos, T., et al. 2000, updated 2002. *AVHRR Polar Pathfinder Twice-Daily 1.25 km EASE-Grid Composites*. Digital Media. Boulder, CO: National Snow and Ice Data Center.
- Scambos, T., et al., 2007. MODIS-based Mosaic of Antarctica (MOA) data sets: continent-wide surface morphology and snow grain size. *Remote Sensing of Environment*, **111**, 242–257.
- Schutz, B., et al., 2005. ICESat mission overview and history. *Geophysical Research Letters*, **32**, L21S01.
- Shepherd, A., and Wingham, D., 2007. Recent sea-level contributions of the Antarctic and Greenland ice sheets. *Science*, **315**, 1529–1532.
- Shi, J. C., and Dozier, J., 1993. Measurements of snow- and glacier-covered areas with single polarization SAR. *Annals of Glaciology*, **17**, 72–76.
- Shi, J. C., and Dozier, J., 1996. Estimation of snow water equivalence using SISR-C/X-SAR. In *Proceedings IGARRS 96*, IEEE No. 96, Chap. 35875, pp. 2002–2004.
- Shi, J. C., and Dozier, J., 1997. Mapping seasonal snow with SIR-C/X-SAR in mountainous areas. *Remote Sensing of Environment*, **59**, 294–307.
- Shi, J. C., Dozier, J., and Rott, H., 1994. Snow mapping in alpine regions with synthetic aperture radar. *IEEE Transactions on Geoscience and Remote Sensing*, **32**, 152–158.
- Sievers, J., Grindel, A., and Meier, W., 1989. Digital satellite image mapping of Antarctica. *Polarforschung*, **59**, 23–33.

- Smith, D. M., 1998. Observation of perennial Arctic sea ice melt and freeze-up using passive microwave data. *Journal of Geophysical Research*, **103**(C12), 27,753–769.
- Smith, N. V., Saatchi, S. S., and Randerson, J. T., 2004. Trends in high northern latitudes soil freeze and thaw cycles from 1988 to 2002. *Journal of Geophysical Research*, **109**, D12101.
- Spreen, G., Kaleschke, L., and Heygster, G., 2008. Sea ice remote sensing using AMSR-E 89-GHz channels. *Journal of Geophysical Research*, **113**, C02S03, doi:10.1029/2005JC003384.
- Stankov, B., and Gasiewski, A., 2004. *CLPX-Airborne multiband Polarimetric Scanning Radiometer (PSR) Imagery*. Digital Media. Boulder, CO: National Snow and Ice Data Center.
- Stroeve, J., and Nolin, A. W., 2002. New methods to infer snow albedo from the MISR instrument with applications to the Greenland ice sheet. *IEEE Transactions Geoscience and Remote Sensing*, **40**, 1616–1625.
- Stroeve, J., Nolin, A., and Steffen, K., 1997. Comparison of AVHRR-derived and in situ surface albedo over the Greenland ice sheet. *Remote Sensing of Environment*, **62**, 262–276.
- Stroeve, J., Markus, T., and Maslanik, J., 2001. Sensitivity analysis of operational passive microwave sea-ice algorithms. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARRS)*, 2001, pp. 1798–1800.
- Strozzi, T., et al., 2008. Estimation of Arctic glacier motion with satellite L-band SAR data. *Remote Sensing of Environment*, **112**(3), 636–645.
- Stuart, K., et al., 2007. The Antarctic iceberg tracking database. <http://www.scp.byu.edu/data/iceberg/database1.html>
- Surazakov, A. B., et al., 2007. Glacier changes in the Siberian Altai Mountains, Ob River basin, (1952–2006) estimated with high resolution satellite imagery. *Environmental Research Letters*, **2**, 045017, 7 pp.
- Tait, A., 1998. Estimation of snow water equivalent using passive = microwave radiation data. *Remote Sensing of the Environment*, **64**, 286–291.
- Tedesco, M., et al., 2005. Analysis of multiscale radiometric data collected during the Cold Land Processes Experiment-1 (CLPX-1). *Geophysical Research Letters*, **32**, L18501.
- Tedesco, M., et al., 2006. Brightness temperatures of snow melting/refreezing cycles: observations and modeling using a multilayer dense medium theory-based model. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(12), 3563–3572.
- Vinje, T., Nordlund, N., and Kvambekk, A. 1998. Monitoring ice thickness in Fram Strait. *Journal of Geophysical Research*, **103** (C5), doi:10.1029/97JC03360.
- Wadhams, P., and Doble, M. J., 2008. Digital terrain mapping of the underside of sea ice from a small AUV. *Geophysical Research Letters*, **35**, L01501, doi:10.1029/2007GL031921.
- Walker, A. E., and Davey, M. R., 1993. Observation of Great Slave Lake ice freeze-up and break-up processes using passive microwave satellite data. In *Proceedings of the 16th Canadian Symposium on Remote Sensing*, Sherbrooke, QC, pp. 233–238.
- Way, J. B., et al., 1997. Winter and spring thaw as observed with imaging radar at BOREAS. *Journal of Geophysical Research*, **102**(29), 671–684.
- Williams Jr., R. S., and Ferrigno, J. G. (eds.), 1988. *Satellite Image Atlas of Glaciers of the World*, Antarctica. U.S. Geological Survey Professional Paper 1386-B, 288 pp.
- Williams, R. S., Jr., et al., 1982. Landsat images and mosaics of Antarctica for mapping and glaciological studies. *Annals of Glaciology*, **3**, 321–326.
- Worby, A., et al., 1999. On the use of electromagnetic induction sounding to determine winter and spring sea ice thickness in the Antarctic. *Cold Regions Science and Technology*, **29**, 49–58.
- Wynne, R. H., et al., 1998. Satellite monitoring of lake ice breakup on the Laurentian shield (1980–1994). *Photogrammetric Engineering and Remote Sensing*, **64**, 607–617.
- Zhang, T., and Armstrong, R. L., 2001. Soil freeze/thaw cycles over snow-free land detected by passive microwave remote sensing. *Geophysical Research Letters*, **28**(5), 763–766.
- Zhang, T., and Armstrong, R., 2003 updated 2005. *Arctic Soil Freeze/Thaw Status from SMMR and SSM/I, Version 2*. Digital Media. Boulder, CO: National Snow and Ice Data Center/World Data Center for Glaciology.
- Zhang, T., Armstrong, R. L., and Smith, J., 2003. Investigation of the near-surface soil freeze/thaw cycle in the contiguous United States: algorithm development and validation. *Journal of Geophysical Research*, **108**(D22), 8860, doi:10.1029/2003JD003530.
- Zhang, T., Barry, R. G., and Armstrong, R. L., 2004. Application of satellite remote sensing techniques to frozen ground studies. *Polar Geography*, **28**(3), 163–196.
- Zhao, Y., Liu, A. K., and Long, D. G., 2002. Validation of sea ice motion from QuikSCAT with those from SSM/I and buoy. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 1241–1246.
- Zhou, G., and Jezek, K., 2002. *DISP Yearly Satellite Photographic Mosaics of Greenland 1962–1963*. Digital Media. Boulder, CO: National Snow and Ice Data Center.
- Zuerndorfer, B., and England, A. W., 1992. Radio brightness decision criteria for frozen ground boundaries. *IEEE Transactions on Geoscience and Remote Sensing*, **30L**, 89–102.
- Zwally, H. J., et al., 1983. *Antarctic Sea Ice, 1973–1976; Satellite Passive-Microwave Observations*. Washington, DC: NASA, SP 459, 206 pp.
- Zwally, H. J., et al., 2008. ICESat measurements of sea ice freeboard and estimates of sea ice thickness in the Weddell Sea. *Journal of Geophysical Research*, **113**, C02S15, doi:10.1029/2007JC004284.

Cross-references

[Cryosphere and Polar Region Observing System](#)
[Ice Sheets and Ice Volume](#)
[Icebergs](#)
[Polar Ice Dynamics](#)
[Sea Ice Concentration and Extent](#)

D

DATA ACCESS

Ron Weaver

National Snow and Ice Data Center, Cooperative
Institute for Research in Environmental Sciences,
University of Colorado, Boulder, CO, USA

Definition

Data access. A process through which users discover and acquire data. In the context of this encyclopedia, data refers to remote sensing data or ancillary or supportive data needed to utilize remote sensing data. Data access is part of the process of acquiring and managing digital data from remotely sensed sources. These processes are collectively considered part of ground data processing systems.

Historical perspective

Data access goes hand in hand with the launch of research and operational satellites in the 1960s. Access in the early 1970s consisted of provision of data to a limited user base most typically as hardcopy photographs or photographic film. Digital data were delivered on nine-track magnetic tapes, and analysis products were often line printer outputs (Swain and Phillips, 1973).

Data access systems of the late 1970s and early 1980s concentrated data access and analysis on workstations, where the scientist/user interacted with data access/analysis systems, which had direct attached disk (or tape) systems storage (Brown and Miller, 1984). The process consisted of the user logging onto these data systems, selecting and analyzing data as a timeshare user.

By the mid-1980s, the advent of the Compact Disk Read-Only Memory (CD-ROM) accelerated the evolution away from centralized data access and analysis computing systems. The CD-ROM allowed significant amounts of data to be cheaply distributed to users who analyzed these

data on their in-house workstations and in many cases personal computers (Kuykendall and Zion, 1984).

With the spread of the Internet, minicomputers, and subsequently the personal computer, digital data access options expanded. Increased bandwidth of the Internet allowed transfer of data electronically. The capabilities of workstations and personal computers allowed the individual research to perform more of their analysis on their local equipment.

Today data access is provided by a complex array of distributed data systems. Use of physical media in data distribution is greatly diminished. Within the data management systems of satellite data centers, data are increasingly housed on rotating disk, with users accessing these data by simple file transfer protocol (FTP) or a more complex array of data discovery, browse, and acquisition tools, many operating in a near-real-time environment. The role of tape media has become one of data security and backup rather than distribution.

Generalized functions of present-day data access systems

Data access per se is the end of a process that starts with data acquisition and ends with delivery of data to an end user or the end user's accessing computer system. Current data systems perform several functions, which typically reside between satellite ground systems and end user analysis systems. They consist of three components: ingest processes, data archival and management services, and finally data discovery and access/distribution services. Note that many systems also include data level change and geophysical data production services. Current day systems may also be distributed or virtual in design and implementation. This means that elements such as ingest or data management are geographically distributed but coupled by networks.

Ingest processing

Data are ingested from a ground system or a science data processing system. The ingest process guarantees that the data received is an exact copy of that transmitted from the antecedent provider and notifies the provider of successful ingest. Typically, data are inserted into online or tape archives with metadata placed in a database-driven inventory system.

Data management

The data management system must safely store the ingested data and provide metadata, allowing discovery and access and management over years to decades. Current systems store data in RAID archives, on magnetic tapes, or optical media. Since the mean lifetime of data storage systems is typically 3–5 years, long-term archival of data means that these data must be migrated from one media/storage environment to a newer system on a regular schedule.

Data discovery and access

Many interfaces between data and users exist, but the fundamental elements include processes to match the user query to the stored data attributes (e.g., data type, geographic location, time span, sensor or data set characteristic) and then delivery of appropriate data. Data visualization and data browse are often employed in data discovery interfaces. These techniques and services allow the user to assess data through visual display rather than just machine-readable digital formats. The discovery and access environment of modern [data access](#) systems is probably the most dynamically changing of any of the three components. As users develop new analysis techniques and tools, the data discovery and access system must adapt to efficiently feed data to these evolving user environments.

Trends in data access

The overall trend is to expose data through a rich variety of interfaces, tools, and services, which allow an increasingly diverse user community to gain access through interfaces native to their discipline or specific community. The access and delivery are becoming increasingly transparent to the end user, where the data resulting from a query simply appear on the user's desktop. Trends and important activity areas are described below, starting with technologies employed in data storage and management data and then access strategies and tools.

Data information technology systems

More remote sensing data are stored on rotating disk, either in RAID systems or in simple disk arrays. Staging data on Internet-accessible rotating disk allows exposure to all types of tools and services. Sometimes data are stored in distributed arrays of rotating disk. These data grids allow for geographic dispersal of data, which can help with network congestion and backup

security. Distributed data systems tend to increase the management demands of a data archive since tracking data physical location(s) is more complex. Eventually it is expected that digital data storage “clouds” will become commonplace. The implications of online storage of data for access are really two: first, the data are available with decreased latency and in some cases near real time, and second, more sophisticated and interactive visualization, browsing, and access tools and methods are possible, especially across multiple data types and data sets.

Remote sensing data access is moving rapidly toward electronic delivery of all data. While network bandwidth still limits the transfer of large volumes, the employment of geographic and temporal subsetting at the data center before delivery across the network can help to mitigate the bandwidth limitation.

Data access strategies

Increasingly, data are separated into levels generally organized according to the processing level. These data levels typically inform data access priorities, storage options, and delivery methods. For example, Level 0/1 data are delivered by bulk transfer methodologies such as the Unix File Transfer Protocol (FTP) or Secure FTP (SFTP) while Level 2–4 data more typically are accessed through viewers, tools, and services. The higher levels are placed in more responsive data systems so users can gain transparent and rapid access to these data sets.

Metadata, or descriptive information about data (see [Data Archival and Distribution](#)), is the lifeblood of data access. Users would not be able to access data in the Internet world without metadata-based search and discovery systems. The trend toward greater dependency on metadata-based searches is driven by both the increased number and diversity of available data sets and the expanded use of remote sensing data by less traditional earth and life science disciplines. To accommodate these needs, more rigor must be applied to the metadata including in some cases semantic translations of terminology between scientific disciplines.

Online data storage allows data delivery systems of a more complex nature. This capability, coupled with the needs of the earth science user community, is driving [data access](#) toward geographically enabled methods. Most importantly users are utilizing geographic information systems (GIS) from commercial vendors such as ESRI and open software consortiums such as the Open Geospatial Consortium. The use of GIS has pushed data providers to deliver data in geospatially enabled formats such as shapefiles and GeoTIFF.

A related trend is the explosion in the use of three-dimensional earth visualizers. Most notable in this category are the Google Earth, Microsoft Virtual Earth, and to a lesser degree open source solutions such as NASA's World Wind. These tools “drape” geolocated earth referenced data over a virtual globe.

Another trend in data access is the use of data delivery protocols that tie multiple data sets and distributed data centers together through one interface. Typically, these interfaces are designed for a specific user community but allow diverse data to be accessed using tools and metadata semantics familiar to the specific user community. Examples of these types of systems include the Open Data Access Protocol (OpenDAP) and the associated Thematic Realtime Environmental Distributed Data Services (THREDDS) system (OpenDAP, 2013; Unidata, 2013) and NASA's Earth Observing System Clearinghouse (ECHO, 2013). In these systems, the user queries data through a web-based Internet graphical user interface. Data are delivered to the user computer transparently, in some cases without clear knowledge of the source data location.

On the international scale, the *Global Earth Observation System of Systems (GEOSS)* is implementing distributed systems using Service-Oriented Architecture (SOA). SOA is based on the premise that complex systems may be broken into a collection of smaller related components. These smaller elements can be tied together through interfaces that intercommunicate application purpose and process logic. In a SOA environment, the service provider publishes a description of the service in a registry. Users can then access multiple services across the GEOSS and combine them in dynamic and varying ways to access data or produce end products.

This rich set of data sources and access tools provides the end user a broad array of options. While much of the data access and browse process is through human-machine interfaces, another growing trend is the use of machine-to-machine data access and automated workflow managers. These workflow systems allow the end user to assemble a processing system from data acquisition from the data archive to generation of a geophysical product and ultimately analysis of that product.

Locating data access systems

Many countries through various data access systems and protocols collect remote sensing data. These systems change often, and thus it is not practical to provide a listing of such systems here. However, two resources are useful for locating data, data access systems, and members of the international remote sensing community. The Global Change Master Directory (www.gcmd.nasa.gov) and its international partners maintain a catalog-level metadata collection for data sets relevant to the environmental research community. The Committee on Earth Observation Satellites (CEOS, www.ceos.org) is an international organization comprised of member countries that launch and operate civilian earth-observing satellites. The CEOS website maintains links to all participating member websites and their data systems. CEOS is also an active participant in the GEOSS effort. These are good first places to start a search for environmental remote sensing data.

Summary

Access systems are changing rapidly. Increasing availability of online (rotating disk) storage, improved network bandwidth, and software allowing distributed data systems and object-based data set retrieval add flexibility to data access and management systems. Users now access larger volumes of data more rapidly and transparently and with more powerful search and access choices than at any point in the short history of satellite remote sensing.

Bibliography

- Brown, J. W., and Miller, C. L., 1984. The NASA/JPL pilot ocean data system. *Oceans*, **16**, 964–969.
- ECHO, (2013). Earth observing system clearinghouse. Available from www.echo.nasa.gov
- EOS, 1995. In Asrar, G. (ed.), *MTPE EOS Reference Handbook*. NASA Goddard Space Flight Centre, Greenbelt, MD, 270 pp.
- Kuykendall, F., and Zion, P. M., 1984. The pilot ocean data system science workstation. *Oceans*, **16**, 970–974.
- OpenDAP, 2013. Open Source Project For a Network Data Access Protocol. Available from <http://www.opendap.org>
- Swain, P. H., and Phillips, T. L., 1973. The role of computer networks in remote sensing data analysis. In *Laboratory for Applications of Remote Sensing, LARS Symposia*. http://docs.lib.purdue.edu/lars_symp/7
- Unidata, (2013). THREDDS, Thematic Real-time Environmental Distributed Data Services. Available from <http://www.unidata.ucar.edu/projects/THREDDS/>

Cross-references

Data Archival and Distribution
 Data Archives and Repositories
 Global Earth Observation System of Systems (GEOSS)

DATA ARCHIVAL AND DISTRIBUTION

Mark A. Parsons

Center for a Digital Society, Rensselaer Polytechnic Institute, Troy, NY, USA

Synonyms

Data curation; Data management; Data stewardship

Definition

Data Archiving. Formally preserving data and information and making it available for an identified but potentially large and changing group of data consumers or users (derived from the ISO standard Open Archival Information System (OAIS) Reference Model (CCSDS, 2002)). This definition applies to the archiving of any type of data or information whether it is a physical sample, a medieval manuscript, a photograph, or a digital data file. Remote-sensing data, with the exception of some early photography, are almost exclusively digital. Therefore, this entry focuses on digital data archiving.

Data Distribution. Movement of data from an archive to an individual user or system (see “[Data Access](#)”).

Digital Preservation. “The series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value. This encompasses not just technical activities, but also all of the strategic and organisational considerations that relate to the survival and management of digital material” (defined by the Digital Curation Centre (DCC) of the United Kingdom (Pennock, 2006, p. 1)).

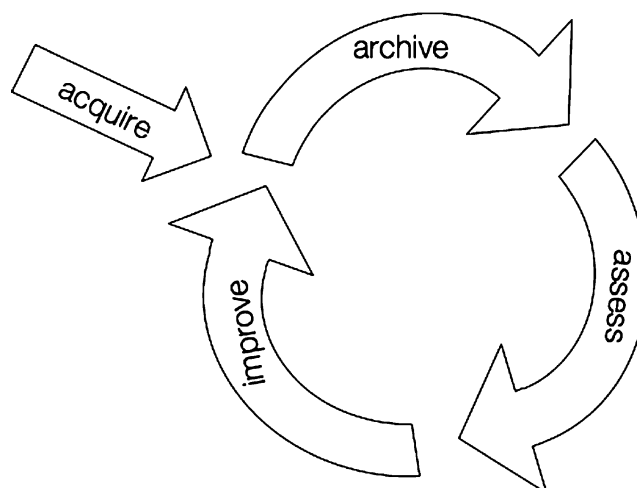
Digital Curation. “Maintaining and adding value to a trusted body of digital information for current and future use; specifically, the active management and appraisal of data over the life cycle of scholarly and scientific materials” (defined by the DCC (<http://www.dcc.ac.uk/about/>)).

Data Management. “Data Resource Management is the development and execution of architectures, policies, practices and procedures that properly manage the full data life cycle needs of an enterprise” (defined by the Data Management Association International (<http://www.dama.org/i4a/pages/index.cfm?pageid=3339>)).

Data Stewardship. “All activities that preserve and improve the information content, accessibility, and usability of data and metadata. These activities include maintaining a scalable and reliable infrastructure to support long-term access and preservation, preserving [data access](#) and archive integrity during media migration and software evolution, providing effective data support services and tools for users, and enhancing data and metadata by adding information that is established throughout the data life cycle” (as defined by the National Research Council Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA (NRC, 2007, p. 41)).

Introduction

Data archiving is an ancient discipline, dating back at least to the third century BC and the Library of Alexandria. Yet now the field is rapidly evolving in response to the new challenges created by the overwhelming growth of digital data. For ages, data archiving meant capturing information on paper in the form of documents, maps, and pictures; cataloging the information; and preserving the paper. Digital data changed all that. While well-preserved paper documents can last and be read for centuries, digital media are much less durable. Consider, e.g., the floppy disk: In a few short decades, it changed size and format several times and now is obsolete. You would be hard-pressed to find a computer that could read an 8 in. floppy disk from the 1970s, and even then, the disk would likely be damaged or the data on it corrupted. Not only do media change, but the data themselves can change in format, presentation, and content; and the volume of data continues to grow. This is especially true in remote sensing, where a growing number of increasingly high-resolution sensors generate terabytes (one terabyte = 1,012 bytes). A typical academic library might contain about two terabytes of printed information of new data every day (NRC, 2007). These data



Data Archival and Distribution, Figure 1 A simple representation of the data life cycle from NRC (2007).

may be in specialized, complex formats that can only be read or analyzed with specialized tools. The data are unique and unrepeatable because they describe features at a moment in time. Yet the data also change as they are calibrated, processed through new algorithms, combined with other data, or corrected for errors.

The dynamic nature of digital data, especially science data, led to the concept of the data life cycle. After data are first created or brought into an archive, users and data managers continually assess and sometimes improve the accuracy and understanding of the data and documentation (metadata). Archivists capture and maintain these improvements for future generations, and the process of acquisition, archiving, evaluation, and improvement continues for as long as the data are considered valuable (Figure 1).

To manage the data through their life cycle, it is insufficient to simply preserve, document, and distribute data. We must actively interact with users in order to curate or steward the data to maintain and increase the value of the data over time. This is reflected in the definitions above and in a growing body of national and international reports. This entry draws substantially from those reports including National Research Council recommendations to NOAA and NASA (NRC, 1998, 2003, 2004, 2007), the International Council for Science’s assessment report on scientific data (ICSU, 2004), the National Science Board’s report on long-lived data (NSB, 2005), and the US Global Climate Research Program’s report – *Global Change Science Requirements for Long-Term Archiving* (USGCRP, 1999).

Data Archiving Responsibilities

The OAIS Reference Model (CCSDS, 2002, pp. 3–1) describes six basic responsibilities of an archive. These responsibilities apply whether you are archiving a medieval manuscript or detailed remote-sensing

imagery, but remote-sensing data have unique considerations because of their volume, complexity, and dynamics. I review here the six OAIS responsibilities in light of data stewardship recommendations from the USGCRP, the NRC, and others to provide an overview of remote-sensing data archiving and distribution. Archiving is an ongoing process, not a onetime action, and the major challenges tend to be organizational and cultural rather than technical.

1. Negotiate for and accept appropriate information from information producers. Stewardship must begin early to ensure adequate planning and collection of information necessary for all archiving responsibilities. Ideally, the data archiving and stewardship process begins before the data are even collected. In remote sensing that might mean as soon as the initial instrument design. Therefore, “negotiation” could include describing how archivists or data scientists will be informed of the instrument design process, as well as determining data flow rates, volumes, formats, etc. Alternatively, archives may need to actively identify and locate data relevant to their community and negotiate for data transfer. Remote-sensing missions may spawn short-term algorithm and product development efforts that need to preserve their products. When archives accept higher-level products, they must also identify the dependencies in the data and ensure that all the related lower-level and ancillary products are also adequately archived and accessible somewhere. In some cases, it may be more appropriate to only archive lower-level products and produce higher-level products on demand. In which case it is especially crucial to track and record algorithm development. An archive may also need to seek out and preserve related ground-based data for validation, calibration, and processing of their remote-sensing data holdings (see “[Calibration and Validation](#)”). Occasionally, “data rescue” efforts may be necessary to acquire neglected, deteriorating, or disorganized data. In all situations, it is necessary to clearly plan roles and responsibilities throughout the full archiving process and to negotiate the cost and duration of archiving.
2. Obtain sufficient control of the information provided to the level needed to ensure long-term preservation. Typically, obtaining “sufficient control” of the data implies ingest or physical transfer of the data into the archive. Increasingly, however, archiving activities may be distributed across multiple organizations. For example, data may physically reside in one location, while archivists develop and maintain access services, user support, and data documentation elsewhere. Nevertheless, an archive must verify data locations and that data are not corrupted during transfer and that all the data in question are transferred. Formalized agreements may be necessary between ground systems, science teams, and archives that verify data flow. Verification techniques may include digital signatures,

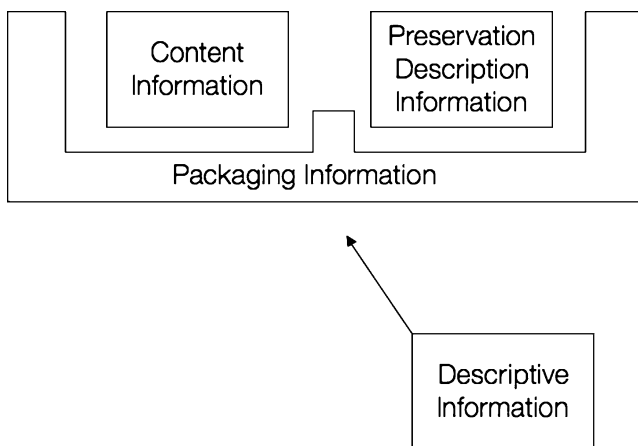
user authentication, and simply checking files against defined listings.

“Sufficient control” may also involve the ability to change the data. The archive may need to conduct quality control processes or reprocess the data with corrected algorithms. The archive will also need to continually revise and enhance the documentation and metadata (information about the data). Archives should work closely with data providers and others when making changes because ownership of digital data can be very complex. Creators of digital data, databases, algorithms, metadata, access tools, and contextual information can all have rights over the materials. Archives need to ensure fair attribution of data authors. (A developing best practice is to formally cite data authors and publishers much as you would a journal article or a book.) They may need to restrict or carefully document data use. They may not even have full control of the data because of privacy or security issues. Managing rights and data control is an ongoing and challenging aspect of data stewardship. See also “[Data Policies](#).”

3. Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community (“Designated Community: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities” (CCSDS, 2002, pp. 1–10)) and, therefore, should be able to understand the information provided.

Defining the designated user community for a data set is a core responsibility of archiving that informs all other responsibilities. The data to be archived, how they are described, how they are presented, and how actively they are curated all depend on the needs of the designated community. Yet user communities change over time as products become better understood and new applications are conceived. Archivists should seek to maximize the use and value of their data by accommodating the broadest possible use including unanticipated use while discouraging scientifically inappropriate use. When developing data documentation, archivists must challenge their assumptions about the knowledge of data users and creators and write in language and detail appropriate to a broad community. They must clearly and explicitly describe data uncertainties and use durable, flexible, and understandable data formats. In short, they must actively engage with their community to understand how it is evolving (Parsons and Duerr, 2005). The designated community must also help determine what data should be archived and what data can be discarded (NRC, 2007).

4. Ensure that the information to be preserved is independently understandable to the Designated Community. In other words, the community should be able to understand the information without needing the assistance of the experts who produced the information. If data are



Data Archival and Distribution, Figure 2 Information package concepts and relationships according to the OAIS Reference Model (CCSDS, 2002).

to be “independently understandable” now and long into the future, we need to maintain and preserve the basic digital data set and a lot of other information describing the data set. The OAIS Reference Model uses the concept of “information packages,” to describe all the information (data and metadata) that needs to be preserved for a given data set. The model describes two core sets of information – “content information” and “preservation description information.” These are bound together in an archive with “packaging information” and made available to catalogs and data discovery tools through “descriptive information” (Figure 2).

Content information includes the actual data, i.e., the primary thing to be archived, and enough “representation information” for the designated community to understand what the data describe (represent) and basically how to use them. For example, representation information for a gridded, satellite-derived, snow cover data set would include detailed information on the data format, the dimensions and geolocation parameters for the grid, the values in the grid cells such as percent snow cover, and the date and time of the values. It would not include information on the sensor, the algorithms used to create the values, how the data relate to other data, etc. So while the content information is independently understandable at a crude level, for the data to be scientifically useful, you need the other core set of information in the package: the preservation description information.

Preservation description information is all the other information necessary to make the data truly useful and to ensure the data are adequately preserved. The information has four categories: Reference, Fixity, Provenance, and Context. Reference information includes things like object identifiers, aliases, file naming conventions, etc. Fixity information describes methods used to ensure there are no undocumented changes in the data

(e.g., digital signatures). These first two categories are mostly the concern of archivists. Provenance and context information, on the other hand, are of interest and are often created by both archivists and users.

Provenance information describes the history of the data set. This includes information on changes in media or transfer of data between institutions, but in remote sensing, it must often include information on types of calibration and validation (see entry *Calibration and Validation*), data processing levels, algorithms applied, etc. One must know precisely what version of data was used in an analysis to reproduce and verify the results. With the large volumes of remote-sensing data, higher-level products are often processed and delivered to users on demand. This creates challenges for both the archivist and user to track precisely which version of data was used for an application or analysis. Maintaining detailed information on sensor and algorithm development is especially critical when considering changes over time in long data records or time series.

Context information describes the background and circumstances that led to the creation of the data and the relationships between the data and other data sets and documents. With remote-sensing data, content information is probably the most complex and voluminous set of documentation in the information package. The USGCRP (1999), p. 32 provides a list of documentation to consider to ensure Earth science data remain scientifically useful. Most of this list can also serve as a description of context information for remote-sensing data. In particular:

- Instrument/sensor characteristics including preflight or preoperational performance measurements (e.g., spectral response, noise characteristics)
- Instrument/sensor calibration data and method
- Processing algorithms and their scientific basis, including complete description of any sampling or mapping algorithm used in creation of the product (e.g., contained in peer-reviewed papers, in some cases supplemented by thematic information introducing the data set or derived product)
- Complete information on any ancillary data or other data sets used in generation or calibration of the data set or derived product
- Processing history including versions of processing source code corresponding to versions of the data set or derived product held in the archive
- Quality assessment information
- Validation record, including identification of validation data sets
- Any changes in instrumentation, controlling agency, surrounding land use, and other factors that could influence the continuity of the record
- A bibliography of pertinent technical notes and articles, including refereed publications reporting on research using the data set
- Information received back from users of the data set or product

Full scientific understanding requires all this information. Yet different disciplines use terminology in different ways, potentially causing inconsistencies and misunderstandings. This is where specifying and adapting to changes in the designated community is most essential. Context information can be very dynamic and requires active curation by archivists in collaboration with users and providers. Furthermore, the complexity of the metadata does not relate directly to the volume of the data set. Smaller data sets, especially highly processed data, can require as much, or more, context information and active curation as large (lower-level) data sets.

The OAIS provides a good conceptual model for an information package, but it does not specify the precise content of the information. The data community has developed metadata standards that have more detail on the precise nature of information to be stored. Historically, the primary metadata content standard for remote-sensing data was the Federal Geographic Data Committee's *Content Standard for Digital Geospatial Metadata* with remote-sensing extensions (FGDC, 1998, 2002). Now, the ISO 19115 standard (ISO, 2006), largely based on the FGDC standards, is increasingly becoming the broader, more accepted standard. The FGDC/ISO standards are very detailed and complex, but they were not constructed according to the OAIS model. As such, they do not explicitly detail the components of an information package. They are particularly well suited to describe context and representation information and some Provenance, but formal practices have not been established to capture all the Preservation Description Information for remote-sensing data. The research library and archives community created the Preservation Metadata: Implementation Strategies (PREMIS) Working Group to develop a core set of implementable, preservation metadata guidelines based on the OAIS Reference Model (PREMIS Editorial Committee, 2008). Work is just beginning to apply PREMIS guidelines to the archiving of remote-sensing data in conjunction with traditional FGDC/ISO standards (see e.g., Duerr et al., (2009)). Keeping track of and implementing relevant standards is an ongoing archival activity.

5. Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original.

This describes the core preservation responsibility of an archive. Archives must, above all else, ensure the safety and integrity of the data in addition to developing and maintaining the necessary preservation description information described above. The large volumes of digital data present major archival challenges, hence the need for formally "documented policies and procedures." Procedures must, of course, be established for routine backups of the data, but the process goes beyond simple backups. It is important to maintain one or more physically separate duplicates of the data set from which a precise copy of the

original can be generated (e.g., files, databases, and schema). Disaster and security management plans should describe this process as well as security mechanisms for protecting the data from environmental hazards (e.g., floods, earthquakes), human attacks (computer viruses, unauthorized access), and technical hazards (changing software, media degradation) that could render the data unusable. A major activity is ensuring the integrity of the media holding the data. For large remote-sensing archives, media refreshment or migration is an ongoing operational activity. In other words, as soon as you finish migrating the entire collection from one medium (e.g., CDs) to another (e.g., tape archive), it may be necessary to start migrating to the next medium (e.g., RAID). Archivists must track media standards, durability, and use by the community. Furthermore, they must ensure that there is no undocumented change to the data during any sort of transfer. This requires the use of tools and procedures that ensure the "fixity" of the data as described in the prescription description information.

Changing data formats present similar challenges as changing media. Barnum and Kerchoff (2000) describe the risks of preserving information in proprietary format (e.g., MS Word). The companies that develop the formats may not stay in business or maintain backward compatibility with older versions. Remote-sensing data are usually in open formats, but much of the related context information may be in proprietary formats. Archives then need to migrate the data to a different format. This can lead to some loss of functionality or information with each migration. Even when the data are stored in a nonproprietary format (e.g., netCDF, HDF, shapefile), the data may not be able to be maintained forever in their original format. Even so-called standard formats evolve with changes in technology. "Self-describing" formats that include basic metadata in the data file are increasingly popular and useful in remote sensing, but these formats often require special tools to read and may not always be backward compatible with earlier versions (Duerr et al., 2004). Format evolution can cause particular problems in the Earth sciences where it is necessary to have a *consistent* long-term data series in order to detect subtle changes over time. These and related issues have led to calls for the establishment of a digital format registry (Abrams and Seaman, 2003; NSF/DELOS Working Group on Digital Archiving and Preservation, 2003). In response, the Harvard University Library has begun to develop the Global Digital Format Registry (<http://www.gdfr.info/index.html>). Another approach is to establish some sort of "universal data format." CCSDS (2008) is working in that direction by creating a standard approach for packaging data and metadata, including software, into a single package called an XML Formatted Data Unit (XFDU) to facilitate information transfer and archiving. Regardless of the approach used, it is important to recognize that while data may be archived in one (relatively stable) format, they may need to be distributed in multiple formats to best meet the needs of different users.

Maintaining processing algorithms is another challenge. It is not always cost effective or practical to preserve all levels of products. Instead an archive may only maintain the lower-level data and process higher-level products as required. This requires ensuring that the algorithms, where the code may be machine dependent, can always be run consistently. This could include maintaining the old computers to run the algorithm, recoding the algorithm to run on newer machines (producing the same results), or emulating an earlier computing environment on a new machine.

Overall, archivists need to carefully consider an appropriate balance of preservation strategies, including media and format *migration*, *normalization* on a few technology-independent standards, and emulation of earlier technologies (NSF and LoC, 2003). Much as with metadata standards, archivists must track developments and standards in media, formats, system architectures, and related technologies to keep their knowledge up to date in a quickly changing environment. This implies that training and formal professional development should also be included as archiving activities (ICSU, 2004).

6. Make the preserved information available to the Designated Community. Traditionally, making data available meant providing a basic catalog of items with controlled access to the materials, i.e., users had to request a copy or borrow the original. Now, with huge and growing volumes of remote-sensing data available through the Internet, [data access](#) has become a much more complex issue involving human interface design, information science, systems engineering, and other fields (see "[Data Access](#)"). Specialized data access mechanisms may be developed for different user communities, but an archive must, at a minimum, maintain a basic catalog of its holdings and distribute the data through standard mechanisms.

Most remote-sensing archives describe their data in directories using the FGDC or ISO metadata standards or some variant. Historically, they would distribute their data on media such as film, magnetic tape, or CD-ROM. Most data are now available through the Internet through standard methods such as FTP but increasingly through more sophisticated protocols such as defined by the Open Geospatial Consortium, the Open-source Project for a Network Data Access Protocol (OPeNDAP), and others. These more advanced protocols often help archives maintain their data in robust formats for preservation while serving the data in formats that work well with the changing tools of their user community. Although most data are distributed through the Internet, large volumes of data still need to be distributed on media. Current media include DVD and different forms of digital tape, but this will continue to change. Ultimately, the increasing volume of remote-sensing data makes distribution of the raw data impractical. Archives are increasingly providing tools that

enable users to subset and process data remotely so they do not need to actually transfer large volumes of data. Furthermore, remote-sensing data are highly distributed, so archives are increasingly working collaboratively to share data, computational power, and services across networks either through grid-computing networks, specialized data systems, or informally shared web services (see "[Data Access](#)," "[Data Archives and Repositories](#)," and "[International Collaboration](#)").

Conclusion

More scientific data have been collected and made available in the past 5 years than were available in all of history before that (Hey, 2006). Much of these data are derived from satellite and airborne remote-sensing instruments. It is vital to manage and preserve those data in accordance with recognized archival procedures, but with scientific data, especially complex remote-sensing data, an archival program must also include proper stewardship of the data. Archivists must actively engage the scientific community in the process to continually document and improve the data. They must develop and adhere to professional principles and best practices. Furthermore, these activities must continue as long as the data have value – potentially forever. This is a challenge for all of science and society. Technical issues present problems, but the largest challenge is creating the organizational and cultural changes necessary to sustain perpetual preservation and new knowledge.

Bibliography

- Abrams, S. L., and Seaman, D., 2003. Towards a global digital format registry. In *World Library and Information Congress: 69th IFLA General Conference and Council*, August 1–9, 2003, Berlin: International Federation of Library Associations and Institutions.
- Barnum, G. D., and Kerchoff, S., 2000. Preserving a tradition of access to United States Government Information. In *2000 Preservation: An International Conference on the Preservation and Long Term Accessibility of Digital Materials*, December 6–8, 2000. York: OCLC.
- CCSDS (Consultative Committee for Space Data Systems), 2002. *Reference Model for an Open Archival Information System (OAIS) CCSDS 650.0-B-1 Issue 1*. Washington, DC: CCSDS Secretariat.
- CCSDS (Consultative Committee for Space Data Systems), 2008. *XML Formatted Data Unit (XFDU) Structure and Construction Rules. Draft Recommended Standard CCSDS 661.0-B-0*. Washington, DC: CCSDS Secretariat.
- Duerr, R., Parsons, M. A., Marquis, M., Dichtl, R., and Mullins, T., 2004. Challenges in long-term data stewardship. In *Proceedings of the 21st IEEE Conference on Mass Storage Systems and Technologies*. IEEE.
- Duerr, R., Parsons, M. A., and Weaver, R., 2009. A new approach to preservation metadata for scientific data—a real world example. In Di, L., and Ramapriyan, H. K. (eds.), *Standard-Based Data and Information Systems for Earth Observation*. Berlin: Springer, pp. 123–142.

- FGDC (Federal Geographic Data Committee), 1998. FGDC-STD-001-1998. *Content standard for digital geospatial metadata* (revised June 1998). Washington, DC: Federal Geographic Data Committee.
- FGDC (Federal Geographic Data Committee), 2002. FGDC-STD-012-2002. *Content standard for digital geospatial metadata: extensions for Remote Sensing Metadata*. Washington, DC: Federal Geographic Data Committee.
- Hey, T., 2006. e-Science and cyberinfrastructure. *Keynote lecture at the 20th International CODATA Conference*, Beijing, China, October 24, 2006.
- ICSU (International Council for Science). 2004. *ICSU Report of the CSPR Assessment Panel on Scientific Data and Information*. Paris: ICSU.
- ISO (International Organization for Standardization), 2006. *ISO 19115 Geographic Information – Metadata*. Paris: ICSU.
- NRC (National Research Council), 1998. *Review of NASA's Distributed Active Archive Centers*. Washington DC: The National Academies Press.
- NRC (National Research Council), 2003. *Government Data Centers: Meeting Increasing Demands*. Washington DC: The National Academies Press.
- NRC (National Research Council), 2004. *Climate Data Records from Environmental Satellites*. Washington DC: The National Academies Press.
- NRC (National Research Council), 2007. *Environmental Data Management at NOAA: Archiving, Stewardship, and Access*. Washington DC: National Academies Press, p. 116.
- NSB (National Science Board), 2005. *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. Washington DC: National Science Foundation, p. 87.
- NSF (National Science Foundation), and LoC (Library of Congress), 2003. *It's About Time: Research Challenges in Digital Archiving and Long-Term Preservation*. Washington DC: Library of Congress.
- NSF/DELOS Working Group on Digital Archiving and Preservation, 2003. *Invest to Save – Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation*, DELOS.
- Parsons, M. A., and Duerr, R., 2005. Designating user communities for scientific data: challenges and solutions. *Data Science Journal*, 4, 31–38.
- Pennock, M., 2006. Digital Preservation – Continued access to authentic digital assets. Digital Curation Center. http://www.jisc.ac.uk/publications/publications/pub_digipreservationbp.aspx.
- PREMIS Editorial Committee, 2008. *PREMIS Data Dictionary for Preservation Metadata version 2.0*. Washington DC: Library of Congress.
- USGCRP, 1999. *Global Change Science Requirements for Long-Term Archiving*. US Global Climate Research Program.

Cross-references

[Calibration and Validation](#)
[Data Access](#)
[Data Archives and Repositories](#)
[Data Policies](#)
[Data Processing, SAR Sensors](#)
[International Collaboration](#)
[Media, Electromagnetic Characteristics](#)
[Policies and Economics](#)
[Remote Sensing, Historical Perspective](#)

DATA ARCHIVES AND REPOSITORIES

Ruth Duerr

National Snow and Ice Data Center, CIRES 449 UCB,
University of Colorado, Boulder, CO, USA

Synonyms

Archive center; Data archive; Data center; Data library;
Data repository

Definitions

Active Archive, Active Archive Center. An organization whose role is to provide processing support, archiving, documentation, and distribution services for information and data during the life of a remote sensing mission, project, or program (derived from NASA's Strategic Evolution of Earth Science Enterprise Data Systems (SEEDS) Formulation Team Final Recommendations Report (SEEDS Formulation Team, 2003)).

Archive Center. Traditionally an archive, archive center, or records center has been defined as “an organization that collects the records of individuals, families, or other organizations; a collecting archives. . . The building (or portion thereof) housing archival collections” where records are “materials created or received by a person, family, or organization in the conduct of their affairs and preserved because of the enduring value. . .” as defined by the Society of American Archivists (Pearce-Moses, 2005). While accurate, this definition concentrates primarily on ingest and preservation activities, making no explicit mention of ensuring that the records preserved are accessible and usable. However, for digital materials such as most remote sensing data and information, accessibility and usability are key concerns and the definition that an archive is “an organization that intends to preserve information for access and use” (CCSDS, 2002) is beginning to be widely adopted.

Data Archive. “A repository which collects databases and data sets.” Defined by the Society of American Archivists (Pearce-Moses, 2005).

Data Center. “A data center (sometimes spelled datacenter) is a centralized repository, either physical or virtual, for the storage, management, and dissemination of data and information organized around a particular body of knowledge or pertaining to a particular business” (Godinho, 2008). Data centers are typically not tied to a single mission or project and may have a long-term charter (SEEDS Formulation Team 2003).

Data Library. An organization that houses a collection of published data. Derived from the library definition of the Society of American Archivists (Pearce-Moses, 2005).

Data Repository. “A place where things [in this case, data] can be stored and maintained; a storehouse. . . . A repository can be a place where multiple databases or files are located for distribution over a network, or

a repository can be a location that is directly accessible to the user. . . .” Defined by the Society of American Archivists (Pearce-Moses, 2005).

Long-Term Archive. “A data archive in which stewardship of data, products, information, and documentation is held on a permanent basis or until the data is considered to be of no value. The stewardship entails preservation, maintenance, and access of the data.” Defined by NASA’s Strategic Evolution of Earth Science Enterprise Data Systems (SEEDS) Formulation Team Final Recommendations Report (SEEDS Formulation Team 2003).

Introduction

Repositories and archives of Earth remote sensing data exist at all scales, from large, permanent organizations sanctioned by governments or intergovernmental organizations to small, ephemeral sites managed by an investigator as a part of their research. As such, it would be both futile and impossible to identify all of the archives scattered around the globe – there are simply too many small-scale, special purpose repositories to enumerate and the archives landscape changes too rapidly. The best that can be done is to categorize the general types of repositories that exist, describe their general characteristics, and discuss general trends. From largest to smallest scope, the range of archives can be divided into international systems of archives, national centers and systems, commercial archives, state and local agencies, and research archives.

International systems

The original system of World Data Centers (WDCs) was set up during the International Geophysical Year (IGY) in 1957–1958 under the sponsorship of the International Council of Scientific Unions (ICSU) to facilitate data archival and distribution, in particular distribution across cold war boundaries. Organized by scientific discipline, both Russian (WDCB’s), American (WDCA’s), and sometimes Western European (WDCC1’s), Asian, or Australian (WDCC2’s) archives were established with the assumption that holdings of each archive would be shared with its disciplinary partners. Funded by their host nation, WDCs provided their data freely to scientists in any country and were as a consequence forbidden from holding classified or confidential data (WDC Panel, 1996). The WDC system’s holdings were broader in scope than Earth remote sensing data, covering a wide range of solar, geophysical, and environmental disciplines. Roughly a dozen of the 52 centers held some Earth remote sensing data, while two WDCs explicitly dealt with remotely sensed land and atmosphere data (NGDC, 2003). In many cases, a WDC was colocated with a national data center specifically tasked and funded to handle the Earth remote sensing data acquired by that country. With the growth of technology and advent of high data rate remote sensing equipment, it proved impractical for these systems to deal with the quantities of data generated. In 2007 ICSU considered how best to “assert

a much-strategic leadership role on behalf of the global scientific community in relation to the policies, management, and stewardship of scientific data and information” (ICSU-SCID, 2007, p. 3). As a result, the former ICSU World Data Centres and former Federation of Astronomical and Geophysical data analysis Services (FAGS) were disbanded and a new World Data System (WDS) was created by the 29th ICSU General Assembly with the goals of

- Enabling universal and equitable access to quality-assured scientific data, data services, products and information,
- Ensuring long term data stewardship,
- Fostering compliance to agreed-upon data standards and conventions, and
- Providing mechanisms to facilitate and improve access to data and data products.

To ensure the authenticity, integrity, confidentiality and availability of the data and services within the WDS, organizations that wish to belong must undergo both initial and ongoing certification processes (ICSU-WDS, 2011). Many of the original WDCs and several of the FAGS, as well as several new organizations have gone through the initial process. At this juncture there are 52 regular members and 7 network partner members of the new WDS federation.

Also tracing their origin to the IGY, the World Meteorological Organization’s (WMO’s) Global Atmosphere Watch (GAW) has a system of five data centers that archive data from a worldwide network of stations. While the scope of data archived by these centers again is broader than Earth remote sensing data, atmospheric profile information provided by a variety of remote sensing instrumentation is archived. A total of five archives exist, organized by measurements including ozone, UV, solar radiation, greenhouse gases, and aerosols. Like many ICSU World Data Centers, the WMO data centers are operated and maintained by their individual host institutions and make data freely available to the scientific community. One difference between the two is an explicit requirement by the WMO DCs that any data used is acknowledged and that coauthorship be offered to the person who acquired and submitted the data to the archive (GAW, 2008).

In addition to the archive systems described above which are dedicated to providing the global science community with access to data, there are a wide variety of international programs and systems that acquire, archive, and distribute Earth remote sensing data as a by-product of their mission, a mission which is often focused on topics of societal concern such as emergency response and early warning of famines. In these cases, the products archived and distributed are often high-level-analysis products rather than raw data, and the long-term accessibility of the data may be uncertain. As an example, the UN-sponsored Global Information and Early Warning System (GIEWS) provides online access to 10 day NDVI-based vegetation anomaly maps for Africa, Asia,

Central and South America, and the Caribbean derived from SPOT imagery going back to 2004 (GIEWS, 2008).

National centers and systems

Over the years, ground-station-based national archives have proliferated. This trend began with the US series of Landsat spacecraft, which continuously transmit data so that any country capable of obtaining a ground system can acquire Landsat data within the mask of their ground station, and has increased dramatically in recent years as the cost of the infrastructure needed has dropped. Access to data from these archives depends not only on the access policy of the country in which the ground station is based but also on the terms of the agreement between that country and the country that owns the spacecraft. The situation may even be further complicated if commercial organizations are involved.

Similarly, due to the decreasing costs of information technology and the development of small, inexpensive spacecraft, the number of countries that own or are participating in a consortium that own a remote sensing satellite is steadily increasing. At this point more than 2 dozen countries, including several third world countries such as Nigeria, have remote sensing satellite systems on orbit or in development (Krebs, 2008). In many cases, development of the programs needed to support space and ground systems is seen as an “essential tool of Socio-Economic development and enhancement of the quality of life of its people” (NASRDA Undated, p. 4).

Irrespective of how the country obtains remote sensing data, its data archives are typically sponsored by one or more government agencies, typically those agencies responsible for space, meteorological information, disaster response, agriculture, or geological surveys. Archives are also often collocated with other entities such as a World Data Center, a ground station, a University, or other research organization.

Commercial archives

Remote sensing as a commercial enterprise can arguably be traced to the early 1920s with the commercial success of aerial photographic maps of Manhattan Island acquired by Sherman Fairchild (PAPA, 2001). Today there are literally hundreds of commercial aerial photographic and remote sensing companies of all sizes scattered around the globe utilizing the full range of airborne platforms including airplanes, helicopters, and balloons to acquire data for a wide range of purposes. Archive policies vary widely from company to company. Particularly for small companies, survival of the archive past the life of the original owner is often not ensured and depends strongly on the fate of the company itself. In some cases, the archive may be transferred to another commercial or government archive as a result. As an example, the estate of William H. Dehn donated the archive of sea ice charts of Alaska, the western Canadian Arctic, and Bering Sea acquired by his company Sea Ice Consultants, Inc. during the years

1953–1986 to the National Snow and Ice Data Center (<http://nsidc.org>) upon his death (NSIDC/WDC for Glaciology, 2005).

Commercial sales of space-based remote sensing data began in the 1980s with the US Landsat system, the French Satellite Probatoire d’Observation de la Terre (SPOT) program, and India’s Indian Remote Sensing Satellite (IRS-1A) series. In each of these cases, the spacecraft itself was government owned, though in the Landsat case the archive was privately marketed. While the Landsat commercialization model itself was not deemed a commercial success, the pattern of commercial sales of data from government-sponsored satellites or acquired by government-funded ground stations continues today throughout much of the world. In fact, a recent study indicates an increasing trend toward “public-private partnerships” for both the space and ground segments of high-resolution remote sensing systems (Gabrynowicz, 2007).

The first license granted to a private company to build and operate its own commercial remote sensing system was granted in 1993 to WorldView Imaging Corporation (now part of DigitalGlobe, <http://www.digitalglobe.com>). Since that year, several other commercial systems have been developed, most of which are targeted at the high-resolution, submeter imaging market where, unlike many science missions, imaging the entire globe especially on a repeating basis is not physically possible and archive growth is typically measured in square kilometers. Such companies, having made a significant investment in the data acquired, often consider themselves content providers, providing robust, redundant, geographically distributed archive systems for business continuity purposes. There are national security implications with the availability of these high-resolution data which one company in particular, ImageSat International, exploits explicitly with its access and pricing programs which offer “a completely autonomous, secret, regional very-high-resolution imaging capability” (<http://www.imagesatintl.com/>). National security concerns are one reason why access to data from these types of high-resolution systems is generally narrowing (Gabrynowicz, 2008). In recognition of the value to many customers of moderate resolution, large-area coverage, with frequent revisits for repetitive monitoring, yet another company, RapidEye has very recently launched a series of five identical spacecraft with 5 m resolution (<http://www.rapideye.com>).

State and local agencies

State and local agencies have been involved in the collection of Earth remote sensing data from the very earliest days, primarily as consumers of high-resolution imagery used for land use management and infrastructure planning. As such, many of the data products in their repositories may have restrictions against third party distribution depending on the terms under which the data were acquired. In many cases, products derived from the original imagery, often in a Geographic Information

System (GIS) format, are available. Such data may not be archived in a standard fashion, and long-term availability is uncertain.

Research data

Increasing amounts of data, including Earth remote sensing data, are being made available by a range of research projects, from individual investigators to larger coordinated observing programs. Typically, the data are available from a project website or via an ftp directory, and increasingly specific program funding is used to develop advanced access capabilities. There are a host of reasons for this trend. The decrease in cost of information technology makes it technically feasible for even small projects, while funding agencies are increasingly requiring that the data assets they paid for be shared (Lynch, 2008). On the other side of the funding equation, investigators are beginning to recognize that having data from their previous work available can help when it is time to get that next grant (Gogenini, 2008). Such data are rarely archived in a standard fashion, and the long-term availability beyond the term of the grant or program that funded the site is uncertain. Moreover, the infrastructure to acquire and formally archive these data for the long-term often does not exist, a situation that may be changing. For example, in the United States the National Science Foundation, through its DataNet program, is attempting to build the kind of sustainable data stewardship capabilities into university systems that would be needed to manage these data for the long term.

Conclusions

The archives and repositories landscape for Earth remote sensing data is undergoing a sea change, from a period where at the small scale there often was no long-term plan for data or data sharing and on the large scale data was held in large monolithic often less than easily accessible archives, toward a period where data at all scales may have a long-term home and be accessible albeit via an extremely diverse and distributed set of facilities. These trends are consistent with general trends in society as a whole, where increasing attention is paid to developing core competencies and other outsourcing activities. For archives this is leading to increasing distribution of functionality, where the organization with the IT infrastructure may not be the organization with the data stewardship knowledge and responsibility and neither may be the organization that provides advanced access services. For example, in the USA, after a contentious start, NOAA's Comprehensive Large Array Data Stewardship System (CLASS) has evolved to comprise solely NOAA's enterprise IT infrastructure in support of the archive missions of the NOAA Data Centers (Casey, 2008).

This trend toward diversification and distribution of functionality is in opposition to a trend toward the kinds of data and service interoperability needed to support both science which is becoming increasingly interdisciplinary

and to resolve societal needs as called for by the 2002 World Summit on Sustainable Development and by the Group of Eight. These high-level meetings called for international collaboration in making Earth observation data available to support nine areas of societal benefit. As a result, 74 governments, the European Commission, and "51 intergovernmental, international, and regional organizations with a mandate in Earth observation or related issues" are currently participating in the development of a Global Earth Observing System of Systems (GEOSS) (http://www.earthobservations.org/about_geo.shtml). Among the intergovernmental organizations participating in GEOSS, the Committee on Earth Observation Satellites (CEOS) is taking the lead in ensuring coordination of satellite-based remote sensing observations globally. If GEOSS is successfully implemented, access to the Earth remote sensing archived anywhere in the world will be available through a single portal.

Bibliography

- Casey, K., 2008. The NEW comprehensive large array data stewardship system (CLASS) – NOAA Data Center Paradigm. Presentation and discussion at the July 16–18, 2008, *ESIP Federation Meeting*, Durham, NH. Available from World Wide Web: http://wiki.esipfed.org/index.php/NOAA_Sessions_on_NIDIS%2C_CLASS%2C_the_Data_Centers%2C_and_much_much_more
- Consultative Committee for Space Data Systems (CCSDS), 2002. Reference Model for an Open Archival Information System (OAIS) CCSDS 650.0-B-1, Issue 1. Washington, DC: CCSDS Secretariat.
- Gabrynowicz, I. "The Land Remote Sensing Laws and Policies of National Governments: A Global Survey." Report for the US Department of Commerce/National Oceanic and Atmospheric Administration's Satellite and Information Service Commercial Remote Sensing Licensing Program of January (2007).
- Gabrynowicz, J. (Director, National Center for Remote Sensing, Air, and Space Law), August 26, 2008 [email].
- GAW (Global Atmosphere Watch), 2008. *World Data Centres*. Switzerland: WMO. Available from World Wide Web: http://www.wmo.int/pages/prog/arep/gaw/world_data_ctr.html. Accessed September 12, 2008.
- GIEWS, 2008. *Global Information and Early Warning System on Food and Agriculture*. Italy: FAO. Available from World Wide Web: <http://www.fao.org/giews/ENGLISH/index.htm>. Accessed September 12, 2008.
- Godinho, R., (2010) "What is data center? – Definition from WhatIs.com" Available from World Wide Web: <http://searchdatacenter.techtarget.com/definition/data-center> <http://searchdatacenter.techtarget.com/definition/data-center>.
- Gogenini, P., August 19, 2008. Comments during seminar question and answer period.
- ICSU-SCID (*Ad Hoc* Strategic Committee on Scientific Information and Data), March 2007. *Draft Final Report to the ICSU Committee on Scientific Planning and Review* (draft). ICSU.
- ICSU-WDS, 2011, "Certification of WDS Members", http://icsuwds.org/images/files/WDS_Certification_Summary_11_June_2012.pdf. Accessed June 19, 2013.
- Krebs, G., 2008. *Gunter's Space Page – Information on Launch vehicles, Satellites, Space Shuttle and Astronautics*. Available from World Wide Web: <http://space.skyrocket.de/>. Accessed August 26, 2008.

- Lynch, C., 2008. Big data: how do your data grow? *Nature*, **455**(7209), 28–29, doi:10.1038/455028a. Available from World Wide Web: <http://dx.doi.org/10.1038/455028a>
- National Geophysical Data Center (NGDC), 2003. *List of World Data Centers by Type*. Boulder: NGDC. Available from World Wide Web: <http://www.ngdc.noaa.gov/wdc/list.shtml>. Accessed August 26, 2008.
- National Space Research and Development Agency (NASRDA), Undated. *National Space Policy and Programmes*. Abuja: Federal Ministry of Science and Technology. Available from World Wide Web: <http://www.nasrda.net/NationalSpaceBook.pdf>. Accessed August 26, 2008.
- NSIDC/WDC for Glaciology, Boulder, Compiler, 2005. *The Dehn Collection of Arctic Sea Ice Charts, 1953–1986*. Boulder, CO: National Snow and Ice Data Center/World Data Center for Glaciology. Digital Media.
- PAPA, 2001. *Professional Aerial Photographers Association (PAPA) International: History of Aerial Photography*. Available from World Wide Web: <http://www.papainternational.org/history.html>. Accessed August 26, 2008.
- Pearce-Moses, R., 2005. *A Glossary of Archival and Records Terminology*. Chicago, IL: Society of American Archivists. Available from World Wide Web: <http://www.archivists.org/glossary/>. Accessed August 26, 2008.
- SEEDS Formulation Team, 2003. *Strategic Evolution of Earth Science Enterprise Data Systems (SEEDS) Formulation Team Final Recommendations Report*. NASA. Available from World Wide Web: <http://eos.nasa.gov/seeds/>
- WDC Panel (ICSU Panel on World Data Centres), April 1996. *Guide to the World Data Center System*. Boulder, CO: World Data Center A. Available from World Wide Web: <http://www.ngdc.noaa.gov/wdc/guide/wdcguide.html>.

Cross-references

Commercial Remote Sensing
 Data Access
 Data Archival and Distribution
 Data Policies
 Global Earth Observation System of Systems (GEOSS)
 Law of Remote Sensing
 Policies and Economics
 Public-Private Partnerships
 Remote Sensing, Historical Perspective

DATA ASSIMILATION

Dennis McLaughlin
 Department of Civil and Environmental Engineering,
 Massachusetts Institute of Technology, Cambridge,
 MA, USA

Synonyms

Inverse problems; Model calibration; Objective analysis;
 Parameter estimation; State estimation

Definition

Data assimilation is the process of characterizing a natural system by combining information from numerical models and observations. Such a characterization can have many purposes, including development of improved forecasts

and better process understanding, filling of data gaps, and design of environmental monitoring strategies. Depending on one's perspective, data assimilation either uses data to constrain a numerical model or uses a numerical model to extend and enhance data. Data assimilation applications are typically distinguished from classical parameter estimation applications by (1) an emphasis on spatially distributed dynamic systems (often described by partial differential equations) that yield large numbers of unknowns when discretized and (2) the need to merge large amounts of diverse data (e.g., obtained from instruments with differing accuracy, resolution, and coverage). The large size of most data assimilation problems makes computational efficiency an important design consideration.

History

Contemporary data assimilation descends from a number of different sources, most notably [weather forecasting](#) and classical least-squares estimation. Meteorological data assimilation developed more or less in parallel with numerical weather prediction and the increased availability of new data sources, including remote sensing measurements from airborne and satellite platforms. Meteorologists were among the first to appreciate the need to merge observations and models in a systematic way. This need motivated the development of a variety of objective analysis (or data assimilation) methods that relied on and extended traditional weather forecasting practices (Daley, 1991; Kalnay, 2003).

The traditional focus in meteorological data assimilation was on short- and medium-term forecasting. Now some operational weather forecasting centers also issue retrospective estimates (or “reanalyses”) of past conditions, which are increasingly being used to examine the effects of weather on ocean and land surface processes, chemical cycles, and ecosystems. Examples of reanalysis data include products issued by the US National Centers for Environmental Prediction (NCEP) (Kalnay et al., 1996; Kistler et al., 2001) and the European Centre for Medium-Range Weather Forecasts (ECMWF) (Uppala et al., 2005).

Oceanographers have also developed and applied data assimilation methods for forecasting (Hodur, 1997) and reanalysis (Wunsch and Heimbach, 2007). More recently, hydrologists, climatologists, atmospheric chemists, ecologists, geophysicists, and petroleum engineers have become interested in data assimilation, largely reflecting the availability of new modeling and observation capabilities in their fields (Kasibhatla et al., 2000; Lenters et al., 2000; McLaughlin, 2002; Trudinger et al., 2007; Oliver et al., 2008).

Methods

There are many ways to combine models and data – some quite general and others rather application specific. In most cases, data assimilation methods characterize

a system by generating from measurements point estimates of uncertain model variables. Point estimates are single values that are typically interpreted as “representative” or “most likely.” Many data assimilation problems in meteorology and geophysics are particularly concerned with point estimates of time-invariant parameters such as initial conditions and physical properties. In numerical weather prediction, the focus is usually on estimation of the initial values of the states (or prognostic variables) in the atmospheric equations of motion. The measurements used in meteorological data assimilation include ground-based weather data, radiosonde soundings, and electromagnetic measurements obtained from satellite sensors operating over a range of wavelengths and resolutions. By contrast, in solid earth geophysics and petroleum engineering, the focus is typically on estimation of uncertain rock properties, rather than initial conditions. Relevant measurements include borehole logs, oil and water production observations, and remotely sensed passive and active seismic data.

Point estimates of uncertain time-invariant model parameters are typically obtained by adjusting these parameters to obtain predictions that are as close as possible to measurements, subject to various constraints. This process may be viewed as a way to solve the inverse of the classical forward modeling problem (Banks and Kunisch, 1989). That is, the objective is to derive the model’s inputs from measurements of its outputs rather than to derive the model’s outputs from its inputs.

In many parameter estimation applications, the model structure is assumed to be perfect (or, equivalently, the model equations are imposed as strong constraints in the fitting process), but the measurements are assumed to be imperfect (i.e., the model is not expected to fit the data perfectly). Data assimilation methods that focus on point parameter estimates typically rely on statistical estimation techniques such as maximum likelihood and nonlinear least squares (Lewis et al., 2006). Depending on the context, the method may or may not make explicit probabilistic assumptions about uncertain parameters and/or measurement errors (McLaughlin et al., 2005). Most of the point parameter estimation techniques used in data assimilation involve optimization (e.g., maximization of an appropriate performance index). The optimization methods encountered in practice vary greatly depending on application but generally must be able to handle large nonlinear problems.

It is possible to generalize the parameter estimation process to account for imperfections in both the numerical model and the measurements. One option is to include dynamic (time-dependent) errors in the model state equations. These model errors (and possibly other uncertain model parameters) may be estimated from measurements. When a data assimilation algorithm provides for model errors, the estimated states are usually not the same as the predictions obtained from

an error-free model. Then the data assimilation process can truly be said to balance (or trade off) model and measurement information. The particular balance obtained depends on the relative confidence placed in the model and measurements. There are various probabilistic and deterministic methods for quantifying confidence. Data assimilation techniques that are able to account for model error include recursive (e.g., Kalman) filters and smoothers, expectation-maximization procedures, representer methods, and weak-constraint variational algorithms (Jazwinski, 1973; Moon, 1996; Bennett, 2002; Evensen, 2007).

In remote sensing applications, it is useful to distinguish between data assimilation methods, which typically rely on spatially distributed dynamic models derived from conservation principles, and retrieval methods, which typically rely on correlations between remotely sensed measurements and “ground truth” observations of the uncertain system states (Deepak et al., 1985). Retrieval methods are sometimes used to generate surrogate measurements for data assimilation algorithms. For example, surface soil moisture estimates obtained from a passive microwave radiobrightness retrieval algorithm can be treated as if they were observations in a land surface data assimilation procedure for estimating subsurface soil moisture profiles. An alternative to using retrieval outputs as surrogate measurements is to explicitly model the physical relationship between unobserved states (e.g., subsurface soil moisture) and the actual measurements (e.g., radiobrightness). The advantage of such a measurement model is that it may be applicable outside the range of conditions used to derive the retrieval algorithm. The relative merits of retrieval algorithms and data assimilation algorithms that explicitly model the measurement process are application dependent.

Interest is rapidly increasing in data assimilation methods that provide interval rather than point estimates for uncertain parameters and states. Interval estimates specify the range (and perhaps the probability) of possible values as well as a typical or most likely value. Such estimates usually require more effort to compute than point estimates, but they can be valuable in forecasting applications where it is important to assess predictability, evaluate the likelihood of extreme events, or bound the range of possible outcomes (Palmer and Hagedorn, 2006). A convenient way to obtain an interval estimate that incorporates measurement information is to use a Bayesian approach to update a prior ensemble (or population) of possible parameters and/or states. Methods for accomplishing this include various versions of the ensemble Kalman filter and smoother, particle filters, and Markov Chain Monte Carlo methods (Brooks, 1998; Arulampalam et al., 2002; Evensen, 2007). Most of these methods can account for both measurement and model errors. Ensemble-based interval estimation techniques are generally less used in practice than traditional point estimation techniques, but the pace of

methodological advances suggests that they will become more common in the near future.

Applications

Applications of data assimilation are widely documented in the meteorological and oceanographic literature and, to a lesser extent, in the literature of other fields. Meeting sessions and workshops devoted to data assimilation methods and applications are routinely sponsored by organizations such as the American Meteorological Society, the European Geological Society, the American Geophysical Union, the Society of Industrial and Applied Mathematics, and the Society of Petroleum Engineers. Most practical applications of data assimilation technology can be roughly divided into operational forecasting, primarily for predicting weather and ocean conditions, and reanalysis. The publication lists on the websites of the European Centre for Medium-Range Weather Forecasts (ECMWF, <http://www.ecmwf.int/>) and the National Center for Atmospheric Research (NCAR, <http://www.ncar.ucar.edu/>) provide convenient places to find typical examples of forecasting and reanalysis applications in meteorology. Some examples of oceanographic data assimilation applications are provided in the publication list found on the Estimating the Circulation and Climate of the Ocean (ECCO) program website (<http://www.ecco-group.org/>).

Data assimilation is becoming an adjunct to research in many fields, particularly when it is able to provide credible extensions to sparse data sets. A typical example is the use of global precipitation, wind, and air temperature reanalysis products issued by operational forecasting centers. These data products have been used to study topics as diverse as water supply, climate change, global chemical budgets, ecological cycles, and commodity pricing (Uppala et al., 2005). In fields such as pollution control and enhanced oil recovery, interest in data assimilation is motivated largely by the need for better information to guide real-time control decisions. Data assimilation appears to offer tangible benefits for such applications, but there have as yet been few large-scale demonstrations. It is likely that future data assimilation research will put more emphasis on decision-making and control. This will require the decision-making process to be explicitly included in the data assimilation problem formulation.

Conclusions

Data assimilation is a natural outgrowth of the classical scientific practice of trying to reconcile models and data. This is inevitably an iterative process which requires continual improvements in the models as well as continual efforts to collect new and more informative data. The analytical tools of data assimilation are intended to support and guide the broader effort to improve our understanding of natural systems.

Some of the factors that limit wider application of data assimilation methods include (1) fundamental uncertainties about the physical, chemical, and biological phenomena that are being modeled and observed; (2) the complexity and computational demands of the models and methods that need to be implemented in operational settings; (3) the dependence on simplifying assumptions that may be difficult to defend in practical situations; (4) the difficulty of insuring the quality of the large amounts of data that need to be assimilated; and (5) the need for better ways to distribute, document, and revise assimilation products. Current research is addressing most of these problems, but the rate of progress that can be expected in any particular application area is difficult to predict.

The factors that contribute most to the advancement of data assimilation practice are probably new instruments, better data sets, and fundamental improvements in modeling capabilities. But there is also room for advances in data assimilation methodology. Particularly promising directions for methodological research include (1) application of machine learning, image processing, and model reduction methods that address issues related to problem size and computational complexity; (2) robust nonlinear estimation techniques that are not restricted by traditional assumptions; (3) geometrical estimation methods that focus on the characterization and prediction of spatial patterns; (4) advanced pattern recognition and anomaly detection algorithms that can be used to automate the quality control process; and (5) web-based methods for documenting, displaying, and disseminating data assimilation products. Overall, there is good reason to expect data assimilation methods to become more powerful and more widely used as modeling, measurement, and computational technology continue to advance.

Bibliography

- Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, **50**, 174–188.
- Banks, H. T., and Kunisch, K., 1989. *Estimation Techniques for Distributed Parameter Systems*. Boston: Birkhauser.
- Bennett, A. F., 2002. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge: Cambridge University Press.
- Brooks, S., 1998. Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society (Series D): The Statistician*, **47**(1), 69–100.
- Daley, R., 1991. *Atmospheric Data Analysis*. New York: Cambridge University Press.
- Deepak, A., Fleming, H., and Chahine, M. (eds.), 1985. *Advances in Remote Sensing Retrieval Methods*, A. Hampton, VA: Deepak.
- Evensen, G., 2007. *Data Assimilation: The Ensemble Kalman Filter*. Berlin: Springer.
- Hodur, R. M., 1997. The Naval Research Laboratory's coupled ocean/atmosphere mesoscale prediction system (COAMPS). *Monthly Weather Review*, **125**, 1414–1430.
- Jazwinski, A. H., 1973. *Stochastic Processes and Filtering Theory*. San Diego, CA: Academic.

- Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge: Cambridge University Press.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–471.
- Kasibhatla, P., Heimann, M., Rayner, P., Mahowald, N., Prinn, R. G., and Hartley, D. (eds.), 2000. *Inverse Methods in Global Biogeochemical Cycles*. Washington, DC: American Geophysical Union.
- Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M., 2001. The NCEP NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bulletin of the American Meteorological Society*, **82**, 247–267.
- Lenters, J. D., Coe, M. T., and Foley, J. A., 2000. Surface water balance of the continental United States, 1963–1995: regional evaluation of a terrestrial biosphere model and the NCEP/NCAR reanalysis. *Journal of Geophysical Research*, **105**(17), 22393–22426, doi:10.1029/2000JD900277.
- Lewis, J. M., Lakshmivarahan, S., and Hall, S. D., 2006. *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge: Cambridge University Press.
- McLaughlin, D., 2002. An integrated approach to hydrologic data assimilation: interpolation, smoothing, and forecasting. *Advances in Water Resources*, **25**(8–12), 1275–1286.
- McLaughlin, D., O'Neill, A., Derber, J., and Kamachi, M., 2005. Opportunities for enhanced collaboration within the data assimilation community. *Quarterly Journal of the Royal Meteorological Society*, **131**(613, Pt. C), 3683–3693.
- Moon, T. K., 1996. The expectation-maximization algorithm. *Signal Processing Magazine IEEE*, **13**(6), 47–60. 1053–5888.
- Oliver, D. S., Reynolds, A. C., and Liu, N., 2008. *Inverse Theory for Petroleum Reservoir Characterization and History Matching*. Cambridge: Cambridge University Press.
- Palmer, T., and Hagedorn, R. (eds.), 2006. *Predictability of Weather and Climate*. Cambridge: Cambridge University Press.
- Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., and Nikolova, S., 2007. OptIC project: an intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models. *Journal of Geophysical Research*, **112**, G02027, doi:10.1029/2006JG000367.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, I., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J., 2005. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**, 2961–3012, doi:10.1256/qj.04.176.
- Wunsch, C., and Heimbach, P., 2007. Practical global ocean state estimation. *Physica D*, **230**(1–2), 197–208.

DATA POLICIES

Ray Harris
Department of Geography, University College London,
London, UK

Synonyms

Access principles; Data sharing

Definition

A set of principles governing the conditions under which data are provided to users.

Introduction

Data policies in remote sensing concern the conditions under which data are provided by the organizations that own and operate remote sensing satellites. The topics that are common in data policy cover ownership, intellectual property, standards, licensing, distribution, pricing, and archiving policies (Harris and Browning, 2005). Many of the organizations that own and operate remote sensing satellites are national governments, and so the data policies have dimensions that are political in nature because they are often part of government policy. Where private sector organizations are involved, they are given licenses to operate by national governments, and therefore, government policy still has a part to play. This discussion examines data policies from the perspective of the main organizations that have influenced data policy.

United Nations

In 1986, the United Nations (UN) agreed a set of 15 principles on remote sensing (Jasentuliyana, 1988; von der Dunk, 2002). The 15 principles were a compromise between those nations that saw outer space as a vertical extension of air space that they could control and those nations that wanted free and open access to outer space for the purposes of exploration. Of the 15 principles, it is especially useful to examine two of them. Principle IV strikes the contrast between “the freedom of the exploration and use of outer space on the basis of equality” on the one hand and “respect for the principle of full and permanent sovereignty of all States and peoples over their own wealth and natural resources” on the other. Remote sensing from satellites can provide more information to the satellite owner about the natural resources of a particular state or country (e.g., agriculture, forestry, minerals) than the information that the country has about its own resources. To answer the potential conflict in Principle IV, Principle XII maintains that the State that is sensed has the right of access to remote sensing data of its territory as soon as the data are produced, on a nondiscriminatory basis and on reasonable cost terms (von der Dunk, 2002). This means, for example, that Tunisia has

the right of access to satellite remote sensing data of its own territory, but it cannot prevent the data collection in the first place by satellites orbiting over its territory. By contrast, no other country can have access by right to the data of Tunisia, although other countries can purchase the data on the open market.

At the UNISPACE III conference in Vienna in 1999, a new International Charter on Space and Major Disasters was proposed (Ito, 2011). The idea of the Charter is to provide a single channel of remote sensing [data access](#) in times of natural or man-made disasters. The Charter is activated about 30–40 times per year and provides data free of charge in response to requests following disasters such as the tsunami in the Indian Ocean in 2004, Hurricane Katrina in the USA in 2005, Cyclone Nargis in Burma in 2008, and flooding in Gansu Province, China, in 2012. When a disaster occurs, one of the Charter members initiates a request for data. This request is then sent to the satellite operators so that they can task their satellites to acquire the data and make the data available as soon as possible to the authorized users. SPOT data are the most commonly used data type under the Charter.

World Meteorological Organization

The World Meteorological Organization (WMO) agreed in 1995 its Resolution 40 to provide advice on data access among its 187 member national meteorological services. The main text of Resolution 40 is:

Members [the national meteorological services] shall provide on a free and unrestricted basis essential data and products which are necessary for the provision of services in support of the protection of life and property and the well-being of all nations.

This core of free and unrestricted access does not apply to all data. In Europe, EUMETSAT interprets WMO Resolution 40 as applying to its 6-hourly Meteosat data, its WEFAX data, the Meteosat-derived products, and the data offered through its Meteosat Internet service (EUMETSAT, 2006). An interesting technique that EUMETSAT uses is to encrypt most of its Meteosat data, except for the 6 hourly transmissions that are provided free of charge.

United States

The approach by the United States (USA) on data policy for remote sensing is part of a wider policy on data produced by US federal government agencies. All federally produced data are made available to all users with no restrictions and with no copyright protection. This means that data produced by the federal agencies responsible for remote sensing (e.g., NASA and NOAA) are made available with no copyright restrictions on making and distributing further copies. The cost of the data is at the level termed COFUR – the cost of fulfilling a user request – which is usually the marginal cost of making

a CD of the data. NOAA provides data through online web access and by direct broadcast from its satellites free of charge.

The USA has a legal basis for its remote sensing data archive. Under Public Law 102–555 the National Satellite Land Remote Sensing Data Archive (NSLRSDA) was created to act as a long-term repository of Landsat and other remote sensing data. Those organizations in the USA that wish to dispose of a large archive of remote sensing data must first offer the archive to the NSLRSDA, which may accept the archive or may choose not to take it.

Europe

The European Space Agency (ESA) created a data policy for its Envisat satellite in 1998 (ESA, 1998). This data policy was subsequently made more generic and was applied to the other ESA remote sensing missions. The ESA data policy distinguished between two categories of the use of the data: Category 1 use for research and applications developments that are specifically in support of the Envisat mission objectives, including preparation for future operational use, and category 2 use for all other uses such as operational and commercial use (see entry [Operational Transition](#)). ESA is responsible for the distribution of category 1 use data at a marginal cost and delegates the distribution of category 2 use data to distributors that can both sell the Envisat data and develop their own products and services at market prices. In 2010 ESA revised its Earth observation data policy to distinguish between two groups of data sets: a free data set and a restrained data set. ESA retains some overlap with categories 1 and 2.

The European Union is involved in remote sensing data policy primarily as a user of data in support of its own policy objectives and as one of the leaders of the Global Monitoring for Environment and Security (GMES) initiative. In addition, there are certain directives of the European Union that are applicable to remote sensing [data access](#) to data about the environment created by public authorities (EurLex, 2006a), legal protection of databases (EurLex, 2006b), copyright protection, reuse of public sector information, and the INSPIRE directive concerning spatial data.

Security

During the twenty-first century, the USA has produced satellites capable of collecting remote sensing data at smaller and smaller pixel sizes, in 2012 down to 0.5 m pixel size (see entry [Commercial Remote Sensing](#)). This process has been characterized by formal government approval, for example, the 2003 National Security Presidential Directive which described the licensing of very-high-resolution commercial remote sensing satellites, stating how the US government would access these

data and how the data would fit in with US foreign policy (Anon, 2003).

India and Malaysia have shown considerable sensitivity over remote sensing data and national security. In India, data with a pixel size of less than 1 m are subject to a formal clearance procedure (ISRO, 2011). In Malaysia, data with a pixel size less than 5 m are subject to review by an internal security organization to check for sensitive areas. By contrast Google Earth and other data distribution mechanisms routinely provide very-high-resolution remote sensing data to all users without restriction.

Conclusion

Remote sensing is becoming more formal in its international structures (see entry *International Collaboration*). The Global Earth Observation System of Systems (GEOSS) is providing a mechanism for international coordination of existing remote sensing systems, of which the GMES initiative is the European component (Harris and Browning, 2005). GEOSS has developed a set of data sharing principles that requests full and open exchange of data. The European Space Agency and the European Union have produced a joint space policy that defines the nature of the relationships between the two organizations (Madders and Thiebault, 2007). More and more countries are producing national space laws or formal policies to guide their technical decisions, for example, Germany. As part of these formalization processes, data policies are becoming more explicit in defining how users can gain access to remote sensing data.

Bibliography

- Anon, 2003. *US Commercial Remote Sensing Policy*. National Security Presidential Directive, 27, Washington, DC. <http://www.fas.org/irp/offdocs/nsdp/remsens.html>
- ESA, 1998. *Envisat Data Policy*. Paris: European Space Agency, ESA/PB-EO(97)57 Rev. 3, February 19, 1998.
- EUMETSAT, 2006. *EUMETSAT Data Policy*. Darmstadt, Germany, pp. 28 http://www.eumetsat.int/Home/Main/AboutEUMETSAT/LegalInformation/SP_1228227465576?l=en
- EurLex, (2006a). Directive 2003/4/EC of the European parliament and of the council of 28 January 2003 on public access to environmental information and repealing council directive 90/313/EEC. Available from World Wide Web: <http://eur-lex.europa.eu/LexUriServ/>. Accessed July 20, 2006.
- EurLex, (2006b). Directive 96/9/EC of the European parliament and of the council of 11 March 1996 on the legal protection of databases. Available from World Wide Web: <http://eur-lex.europa.eu/LexUriServ/>. Accessed July 20, 2006.
- Harris, R., and Browning, R., 2005. *Global Monitoring: The Challenges of Access to Data*. London: UCL Press/Cavendish Publishing, p. 229. ISBN 18594 1950X.
- ISRO, 2011. *Remote Sensing Data Policy*. Bangalore, India: Indian Space Research Organisation, p. 3. ISRO RSDP-2011.
- Ito, A., 2011. *Legal Aspects of Satellite Remote Sensing*. Leiden: Martinus Nijhoff Publishers, p. 353. ISBN 978-90-04-19032-0.
- Jasentuliyana, N., 1988. United Nations principles on remote sensing. *Space Policy*, 4(4), 281–284.
- Madders, K., and Thiebault, W., 2007. Carpe diem: Europe must make a genuine space policy now. *Space Policy*, 23(1), 7–12.

Von der Dunk, F., 2002. United Nations principles on remote sensing and the user. In Harris, R. (ed.), *Earth Observation Data Policy and Europe*. Lisse, The Netherlands: A A Balkema, pp. 29–40.

Cross-references

[Commercial Remote Sensing](#)
[International Collaboration](#)
[Operational Transition](#)

DATA PROCESSING, SAR SENSORS

Jakob van Zyl

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

Doppler sharpening; Imaging radars

Definition

Synthetic Aperture Radar (SAR). A coherent radar system that records data in a way that allows later processing of the radar returns into images.

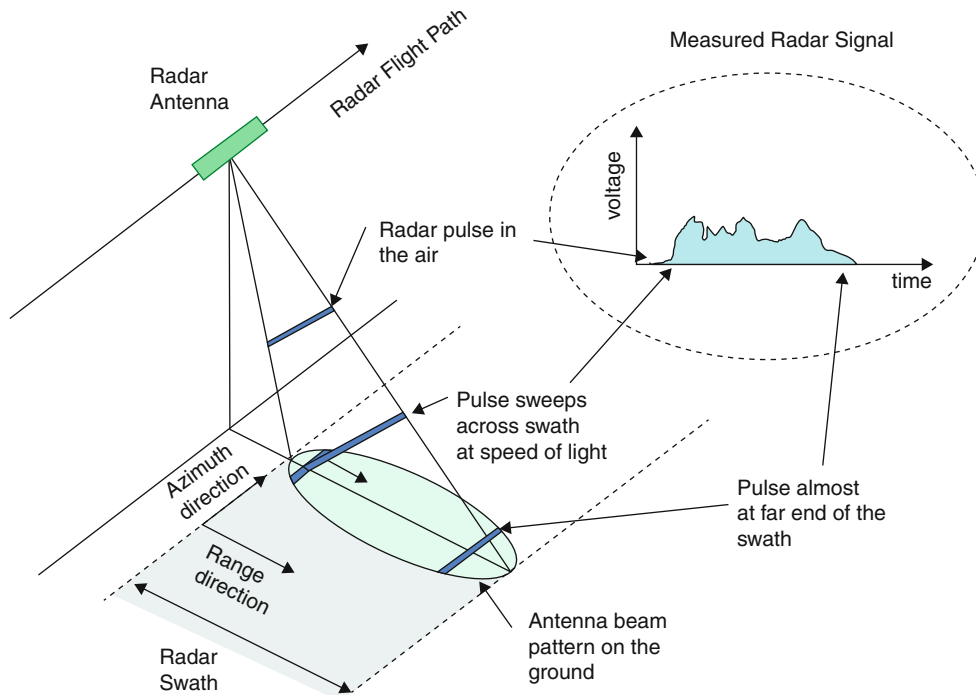
SAR Data Processing. The process by which radar signals are converted into images.

Introduction

The word **RADAR** is an acronym for Radio Detection and Ranging. A radar measures the distance, or *range*, to an object by transmitting an electromagnetic signal and receiving an echo reflected from the object. Since electromagnetic waves propagate at the speed of light, one only has to measure the time it takes the radar signal to propagate to the object and back to calculate the range to the object. The total distance traveled by the signal is twice the distance between the radar and the object, since the signal travels from the radar to the object and then back from the object to the radar after reflection. Therefore, once we measured the propagation time, we can easily calculate the distance to the target.

Radars provide their own signals to detect the presence of objects. Therefore, radars are known as *active* remote sensing instruments. Because the radar provides its own signal, it can operate during day or night. In addition, radar signals typically penetrate clouds and rain, which means that radar images can be acquired not only during day or night but also under (almost) all weather conditions. For this reason, radars are often referred to as *all-weather* instruments. Imaging remote sensing radars such as synthetic aperture radars (SAR) produce high-resolution (from submeter to few tens of meters) images of surfaces.

In this entry, we will describe the process by which radar signals are processed into high-resolution images.



Data Processing, SAR Sensors, Figure 1 Imaging geometry for a side-looking radar system.

Basic principles of radar imaging

Imaging radars generate surface images that are at first glance very similar to the more familiar images produced by instruments that operate in the visible or infrared parts of the electromagnetic spectrum. However, the principle behind the image generation is fundamentally different in the two cases. Visible and infrared sensors use a lens or mirror system to project the radiation from the scene on a “two-dimensional array of detectors” which could be an electronic array or, in earlier remote sensing instruments, a film using chemical processes. This imaging approach conserves the relative angular relationships between objects in the scene and their images in the focal plane. Because of this conservation of angular relationships, the resolution of the images depends on how far away the camera is from the scene it is imaging.

Imaging radars use a quite different mechanism to generate images, with the result that the image characteristics are also quite different from that of visible and infrared images. In order to separate objects in radar images in the cross-track direction and the along-track direction, two different methods are implemented. The *cross-track* direction, also known as the *range* direction in radar imaging, is the direction perpendicular to the direction in which the imaging platform is moving. In this direction, radar echoes are separated using the *time delay* between the echoes that are backscattered from the different surface elements. The *along-track* direction, also known as the *azimuth* direction, is the direction parallel to the movement of the imaging platform. The phase history of the

radar wave is used to separate surface pixels in the along-track dimension in the radar images.

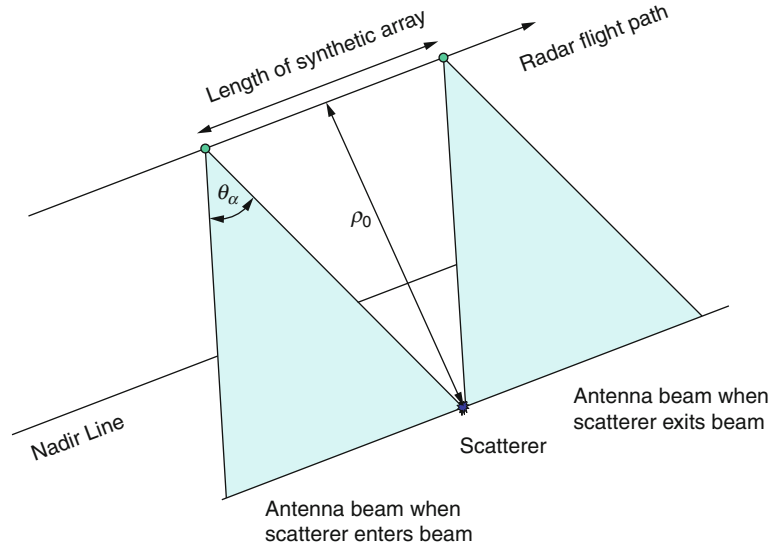
Imaging radar sensors typically use an antenna which illuminates the surface to one side of the flight track. Usually the antenna has a fan beam which illuminates a highly elongated elliptical-shaped area on the surface as shown in Figure 1. The illuminated area across track defines the image *swath*. Within the illumination beam, the radar sensor transmits a very short effective pulse of electromagnetic energy. Echoes from surface points farther away along the cross-track coordinate will be received at proportionally later time (Figure 1). Thus, by dividing the receive time in increments of equal time bins, the surface can be subdivided into a series of *range bins*. As the platform moves, the sets of range bins are covered sequentially, thus allowing strip mapping of the surface line by line.

Radar resolution

The *resolution* of an image is defined as that separation between the two closest features that can still be resolved in the final image. First, consider two point targets that are separated in the line-of-sight (also known as *slant*) range direction by x_r . Because the radar waves propagate at the speed of light, the corresponding echoes will be separated by a time difference Δt equal to

$$\Delta t = 2x_r/c \quad (1)$$

where c is the speed of light and the factor 2 is included to account for the signal round trip propagation as described



Data Processing, SAR Sensors, Figure 2 The synthetic aperture radar integrates the signal from the scatterer for as long as the scatterer remains in the antenna beam.

before. Radar waves are usually not transmitted continuously; instead, radar usually transmits short bursts of energy known as radar *pulses*. The two features can be discriminated if the leading edge of the pulse returned from the second object is received later than the trailing edge of the pulse received from the first feature.

Therefore, the smallest separable time difference in the radar receiver is equal to the effective time length τ of the pulse. Thus, the slant range resolution of a radar is

$$2x_r/c = \tau \Rightarrow x_r = \frac{c\tau}{2} \quad (2)$$

Now let us consider the case of two objects separated by a distance x_g on the ground. The corresponding echoes will be separated by a time difference Δt equal to

$$\Delta t = 2x_r \sin \theta / c \quad (3)$$

The angle θ in Equation 3 is the *local* incidence angle. Therefore, the ground range resolution of the radar is given by

$$2x_r \sin \theta / c = \tau \Rightarrow x_r = \frac{c\tau}{2 \sin \theta} \quad (4)$$

Note that Equation 5 implies that the ground range resolution is different for different incidence angles.

These equations show that in order to produce high-resolution images, a radar system must utilize a short effective pulse length. In practice, this would require the transmission of very-high-power short-duration pulses, something that is quite difficult to do from space. To avoid this, most modern radar systems use *pulse modulation* schemes that allow us to transmit a physically long pulse, which can then be *compressed* during data processing to form a short effective pulse length. The most commonly

used modulation scheme is to vary the radar frequency linearly while the pulse is being transmitted. These signals are known as *chirp* signals, since the frequency change mimics the chirping of a bird. The effective pulse length of a chirp signal is

$$\tau = 1/B \quad (5)$$

where B represents the *bandwidth* of the modulation we used. Note that the resolution depends only on the bandwidth of the system, not how far away the object is.

Synthetic aperture radar

As the radar moves along the flight path, it transmits pulses of energy and records the reflected signals, as shown in Figure 1. When the radar data are processed, the position of the radar platform is taken into account when adding the signals to integrate the energy for the along-track direction. Consider the geometry shown in Figure 2. As the radar moves along the flight path, the distance between the radar and the scatterer changes, with the minimum distance occurring when the scatterer is directly broadside of the radar platform. The *phase* of the radar signal is given by $-4\pi R/\lambda$. The changing distance between the radar and the scatterer means that after range compression, the phase of the signal will be different for the different positions along the flight path.

The range between the radar and the scatterer as a function of position along the flight path is given by

$$R(s) = \sqrt{\rho_0^2 + v^2 s^2} \quad (6)$$

where ρ_0 is the range at closest approach to the scatterer, v is the velocity of the radar platform, and s is the time along the flight path (so-called *slow time*) with zero time at the

time of closest approach. To a good approximation for remote sensing radars, we can assume that $v s \ll \rho_0$. In this case, we can approximate the range as a function of slow time by a Taylor series expansion of the square root as

$$R(s) \approx \rho_0 + \frac{v^2}{2\rho_0} s^2 \quad (7)$$

The phase of the signal then becomes

$$\phi(s) = -\frac{4\pi R(s)}{\lambda} \approx -\frac{4\pi\rho_0}{\lambda} - \frac{2\pi v^2}{\rho_0 \lambda} s^2 \quad (8)$$

The instantaneous frequency of this signal is

$$f(s) = \frac{1}{2\pi} \frac{\partial \phi(s)}{\partial s} = -\frac{2v^2}{\rho_0 \lambda} s \quad (9)$$

This is the expression of a linear frequency chirp, that is, the instantaneous frequency of this signal varies linearly with slow time. To find the bandwidth of this signal, we have to find the maximum time that we can use in the signal integration. This maximum “integration time” is given by the amount of time that the scatterer will be in the antenna beam. For an antenna with a physical length L , the half-power horizontal beamwidth is $\theta_a = \lambda/L$, so that the scatterer at the range of closest approach ρ_0 is illuminated for a time

$$s_{tot} = \frac{\lambda \rho_0}{Lv} \quad (10)$$

Half of this time occurs when the radar is approaching the range of closest approach, and half of it is spent traveling away from the range of closest approach. Therefore, the bandwidth of the signal shown in Equation 9, which is the Doppler bandwidth of the SAR signal, is

$$B_D = \frac{2v}{L} \quad (11)$$

If this signal is filtered using a matched filter, the resulting compressed signal will have a width in time of $1/B_D$. Since the radar platform moves at a speed of v , this leads to an along-track resolution of

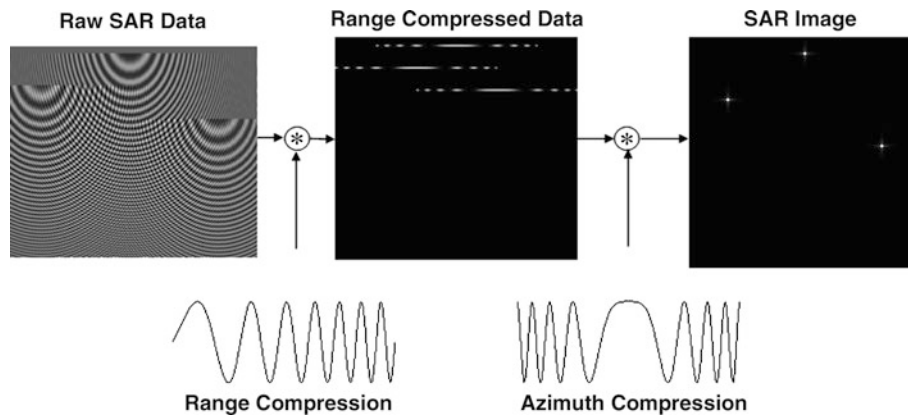
$$x_a = \frac{v}{B_D} = \frac{L}{2} \quad (12)$$

This result shows that the azimuth (or along-track) surface resolution for a synthetic aperture radar is equal to half the size of the physical antenna and is *independent of the distance between the sensor and the surface!* At first glance this result seems most unusual. It shows that a smaller antenna gives better resolution. This can be explained in the following way. The smaller the physical antenna is, the larger its footprint. This allows a longer observation time for each point on the surface, that is, a longer array can be synthesized. This longer synthetic array allows a larger Doppler bandwidth and hence a finer surface resolution. Similarly, if the range between the sensor and surface increases, the physical footprint increases leading to a longer observation time and larger Doppler bandwidth which counterbalances the increase in the range.

SAR processing implementation

Following the description above, we can now illustrate the most commonly used SAR processing algorithm known as the *rectangular imaging* algorithm (Curlander and McDonough, 1991). As the radar moves along its flight path, it transmits pulses of energy at a rate known as the *pulse repetition frequency* (PRF). For each pulse, the received echoes are digitized and stored as a column in a large matrix. Successive columns are recorded for successive pulses. Therefore, the column direction of the data array represents the slant range direction, while the row direction represents the along-track direction.

The SAR processing starts with a matched filter *compression* of the signals in each column of the matrix. This is known as *range compression*. For the mathematic



Data Processing, SAR Sensors, Figure 3 Graphic illustration of the rectangular SAR imaging algorithm. The image contains three point scatterers.

details of this process, see Elachi and van Zyl (2007). This matched filter operation is then repeated in the along-track direction using the appropriate signals with bandwidth given in Equation 10. This step in the processing is known as *azimuth compression*. In between these two steps, there is a step known as *range curvature correction*. This subtle point refers to the fact that as the range changes with slow time (see Equation 7), the echoes from the same point scatterer may not all be in the same row of the matrix. In fact, for spaceborne systems, the signal may extend over many rows as we move in the along-track direction. This is known as *range walk*. In order for us to perform the azimuth compression, we need to use only the signals that come from a particular scatterer. We can do this by predicting the range walk and then by resampling the signals so as to use the phase history along the curved range path.

The rectangular SAR processing process is illustrated graphically in Figure 3. In practice we utilize fast Fourier transforms to perform the matched filter operations.

Summary

SAR processing refers to the process of turning radar signals into images. Here we described the basics behind this imaging process. Many different SAR sensors are providing images on a routine basis today. Most of the research now involves the post-processing of SAR data to extract geophysical information from the images.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Curlander, J. C., and McDonough, R. N., 1991. *Synthetic Aperture Radar System and Signal Processing*. New York: Wiley.
 Elachi, C., and van Zyl, J. J., 2007. *Introduction to the Physics and Techniques of Remote Sensing*. New York: Wiley.

DECISION FUSION, CLASSIFICATION OF MULTISOURCE DATA

Björn Waske¹ and Jón Atli Benediktsson²

¹Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany

²Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland

Definition

In the context of remote sensing-based land cover classification, many applications are aiming on the combination (i.e., fusion) of different data sources, for example, remote sensing images from different Earth Observation systems, in order to improve the classification accuracy. *Decision*

fusion is a potential approach and is defined as the concept of combining information from different data sources, after each source has been classified individually.

Introduction

Over the past decades, numerous Earth Observation (EO) systems have been launched, providing various data sets. On one hand, passive instruments, such as Landsat-5 TM and Landsat-7 ETM+, the ASTER sensor on the Terra platform, Envisat's MERIS, the SPOT satellites, and the AVNIR-2 on ALOS. These well-known optical EO systems are supplemented by passive instruments, such as ERS-2 SAR, Envisat ASAR, and Radarsat-1 as well as more recently launched systems such as TerraSAR-X, ALOS PALSAR, Cosmo-SkyMed, and the enhanced system Radarsat-2.

All these data and instruments have different characteristics, for example, different spatial and spectral resolutions, and diverse repetition rates. Moreover, multispectral and synthetic aperture radar (SAR) instruments operate in different wavelength range of the electromagnetic spectrum, ranging from visible to microwave. Thus, these EO systems provide different but complementary information. The combination of different data sources often improves the classification of remote sensing imagery in terms of the overall accuracy as demonstrated in several studies (e.g., Brisco and Brown, 1995; Benediktsson and Kanellopoulos, 1999; Michelson et al., 2000; Blaes et al., 2005; Waske and Benediktsson, 2007; Koetz et al., 2008).

Besides combining multisensor data sets, for example, multispectral and SAR imagery (Brisco and Brown, 1995; Waske and Benediktsson, 2007) or hyperspectral and LiDAR data (Koetz et al., 2008; Dalponte et al., 2008; Pedergrana et al., 2012), the combination of image time series is promising. Such multitemporal applications are usually more accurate, particularly when dealing with areas that are characterized by temporal variations, for example, due to crop phenology and plant status as in areas dominated by agricultural land use (Brisco and Brown, 1995; Chust et al., 2004; Blaes et al., 2005).

In addition to the use of multisensor images, the use of multisource data sets was discussed in several studies, including spatial information derived from the input images imagery (Benediktsson et al., 2003; Bruzzone and Carlin, 2006; Fauvel et al., 2008; Dalla Mura et al., 2010) and additional auxiliary data, for example, topographical information, statistical data on crop areas, and parcel boundaries or general environmental data (Briem et al., 2002; De Wit and Clevers, 2004; Wright and Gallant, 2007).

Overall, these studies clearly demonstrate that a combination of different data set (i.e., data fusion) is assumed to be valuable in terms of the classification accuracy. Regarding the increased availability of remote sensing imagery as well as auxiliary data, such multisource data

processing approaches become even more attractive. Consequently, the development of appropriate data fusion and classification of multisource data sets is an important ongoing research field of remote sensing (Richards, 2005).

In the following section, data fusion as a term is introduced. Afterward a very brief review of recent developments in land cover classification of remote sensing images is given, followed by a review of selected data fusion applications. Brief conclusions and future outlook are given in the final section.

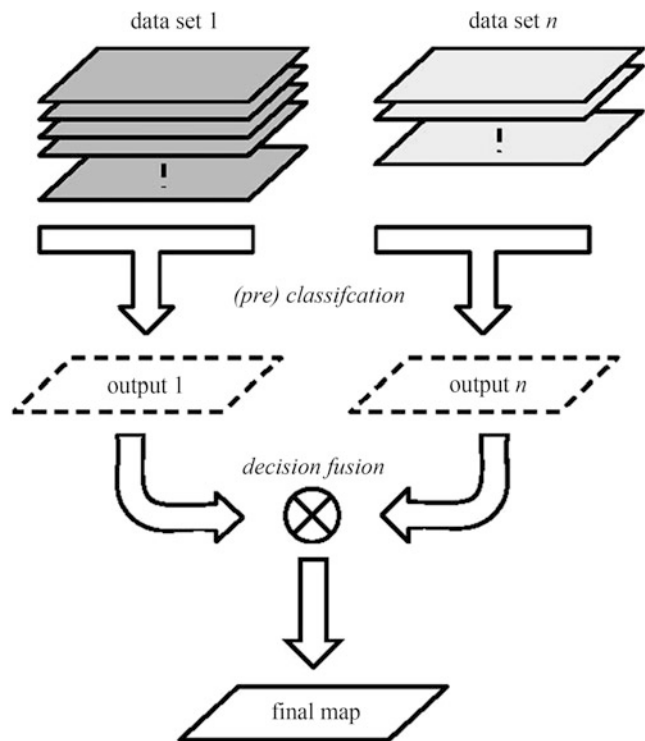
Data fusion: definitions

Different authors have proposed a definition for data fusion. Hall and Llinas (1997) define data fusion as a process which deals with different data sources to achieve more accurate information in the context of decision making. Contrarily, Wald (1999) provides a more specific definition and defines data fusion as a formal concept of tools and means for joining different data sources. The aim of data fusion is generating information of “greater quality,” whereas the definition of “greater quality” depends on the particular application.

However, when dealing with remote sensing applications, data fusion mostly refers to the combination of image data. While the authors mentioned above present general definitions on data fusion, van Genderen and Pohl (1994) focused on image fusion, which was defined as “the combination of two or more different images to form a new image by using a certain algorithm.” In general, the fusion process can be distinguished into three main groups, depending on the level the fusion process is performed (Pohl and van Genderen, 1998):

- *Data-level fusion* refers to the concept of combining the original, raw sensor data on pixel level. The combined data provides the input for the classification algorithm.
- *Feature-level fusion* refers to the combination of different data sources, after a feature vector is extracted from each data set. The features are combined together into a single feature vector, which is used as input for the classifier algorithm.
- *Decision-level fusion* refers to a concept, which is based on the combination of different data sources, after each individual source has been handled separately. In contrast to the original information, the preliminary outputs are combined (e.g., different classification outputs) to generate a final result (see Figure 1).

However, considering the background of this entry, we note that data-level and feature-level fusion only constitute a modification of the available input data, not necessarily with direct consequences on the classification algorithms. Therefore, they are not discussed further here and only a brief review of selected applications is given, whereas decision-level fusion is discussed in more detail. Nevertheless, the discussion of the classifier algorithm itself would be beyond the scope of this entry. A general overview of classification of remote sensing data is given in the entry *Pattern*



Decision Fusion, Classification of Multisource Data,
Figure 1 Schematic diagram of decision fusion.

Recognition and Classification in this book. The interested reader is also referred to one of several textbooks (e.g., Duda et al., 2000; Richards and Jia, 2005) for a detailed introduction of pattern recognition and classification.

Enhanced land cover classification by decision fusion

General developments in remote sensing land cover classification

Recent developments in the classification of remote sensing data have been driven by different factors, such as improved computer power, enhanced Earth Observation systems, and increased requirements on operational applications. While common applications are often based on statistical approaches, which were usually directly taken from the field of signal processing, the increased availability of multisource data sets demands more sophisticated methods (Richards, 2005; Jain et al., 2000). These facts result among others in a change from statistical approaches to more powerful machine learning algorithms.

Nevertheless, well-known statistical classifiers have been used in several multisource remote sensing applications. The conceptually simplest application is based on the stacked vector approach, which uses a composite vector with components from all of the data sources, for example, the digital numbers of each pixel. Although the well-known Gaussian maximum likelihood classifier

was used successfully in this context (e.g., Brisco and Brown, 1995; Chust et al., 2004; Huang et al., 2007), such method is often not adequate when classifying multisource and multitemporal data sets. Different sources can often not be described by a common statistical model, such as a multivariate Gaussian model. Moreover, individual data sources may not be equally reliable. One data set, for example, an image from a specific acquisition date, might be more relevant to classify a specific class, while another image is more applicable to discriminate between other land cover classes. Thus, a weighting of the different data sets could be useful in this context, which statistical techniques usually not allow.

Because of the above, distribution-free or nonparametric techniques are usually more adequate for multisource applications. Neural networks (Benediktsson et al., 1990; Serpico and Roli, 1995), self-learning decision trees (Fitzgerald and Lees, 1994; Friedl and Brodley, 1997), and Support vector machines (SVMs) (Huang et al., 2002; Foody and Mathur, 2004) are some examples of well-known such methods, which have been successfully used in the field of remote sensing. These methods still exhibit further development and adaption, for example, the use of support vector machines in the context of semi-supervised and ill-posed classification problems (Bruzzone et al., 2006; Chi and Bruzzone, 2007) or introducing a kernel framework for multitemporal and multisensor studies (Camps-Valls et al., 2008). Import vector machines (IVMs) are a promising alternative that have been very recently introduced on context of remote sensing. Experimental results show that SVM and IVM perform almost similarly in terms of accuracies, while the probabilities provided by IVM are more reliable, when compared to the probabilistic outputs provided by SVM. This fact is particularly interesting, because probabilities can be used for additional image analysis (Roscher et al., 2012).

Because the above methods usually do not require any prior assumptions about the distribution of input data, as is the case for the maximum likelihood classifier, they are usually better suited for the classification of multisource data sets.

Another development in machine learning and pattern recognition is the concept of multiple classifier systems or classifier ensembles that seems particularly interesting for multisource and high-dimensional remote sensing data sets (Briem et al., 2002; Waske and van der Linden, 2008; Cheung-Wai Chan and Paelinckx, 2008; Waske et al., 2010). In contrast to a regular classification, which is based on one classifier, a set of classifiers is used to generate the final classification results. The approach assumes that different independent classifiers generate individual errors, which are not generated by the majority of the other classifiers (Polikar, 2006).

While many classifier systems combine different variations of the same algorithm, for example, bagging, boosting, and random subspace methods, other applications combine different algorithms (Benediktsson and Kanellopoulos, 1999; Steele, 2000; Waske and van der Linden, 2008). One reason for the latter approach might

be that one method is better to describe one part of the feature space, whereas another algorithm performs better in another part of the feature space (Jain et al., 2000). Thus, these approaches seem particularly interesting for classifying multisource data sets.

Enhanced land cover classification by decision fusion

Several multisource applications are based on decision fusion, an approach which has been used successfully for different purposes and data types (e.g., Jeon and Landgrebe, 1999; Benediktsson and Kanellopoulos, 1999; Bachmann et al., 2003; Fauvel et al., 2006; Gamba et al., 2007; Waske and Benediktsson, 2007). Following the concept of decision fusion, the information from different data sources is combined, after each source has been classified separately.

In Benediktsson and Kanellopoulos (1999), decision fusion was used for the classification of multisensor data (i.e., multispectral and SAR data) as well as a hyperspectral imagery, data set, which was separated into different subsets. Contrary to the multisensor application, which already contains different data sets, the hyperspectral data are divided into several subsets, before the decision fusion is performed. Using the correlation between individual bands, the hyperspectral image is divided into several subsets. Afterward each subset is individually classified by a statistical classifier and a neural network. To generate the final result, the outputs were combined by decision fusion.

Jeon and Landgrebe (1999) introduced the concept for the classification of multitemporal imagery and discussed the use of different fusion strategies. The enhanced classification of very-high-resolution imagery was discussed in Gamba et al. (2007). After separating boundary and non-boundary pixels into two subsets, each subset was classified individually and the final map was generated by decision fusion of the preliminary outputs. Waske and Benediktsson (2007) discussed the use of different algorithms and approaches, including decision fusion, for the classification of multisource data set from an agricultural area. The results clearly demonstrate that the decision fusion outperforms other classification techniques in terms of accuracy.

Although these studies are based in general on the concept of decision fusion, the strategies, which were used for the fusion process, differ. Several methods have been introduced (Kuncheva, 2004), depending among others on the derived output of the classifiers (i.e., class labels and class probabilities, respectively). These methods comprise, for example, among others, a simple majority vote for class labels, average and weighted average rules for class probabilities, as well as more sophisticated approaches as fuzzy integral, Dempster-Shafer theory, and consensus theory (see Kuncheva, 2004 for a review).

The fusion may be based on consensus theory, for example, that employs single probability functions to summarize estimates from various classifiers by

consensus rules (Benediktsson and Swain, 1992). The linear opinion pool and the logarithmic opinion pool are common consensus rules, which are based on the weighted sum and the weighted products of the source-specific a posteriori probabilities. Thus, these approaches require the costly selection of weights, which generally reflect the goodness of the input data. Various selection schemes have been proposed (Benediktsson et al., 1997); nevertheless, the approach still remains computationally relatively complex.

Applications that use simple voting strategies like majority voting and complete agreement are both computationally less complex and able to provide promising results in terms of the classification accuracy. For example, in Benediktsson and Kanellopoulos (1999), a statistical classifier and a neural network were trained parallelly on the same data set. The different sources of a multisensor data set, consisting of multispectral and SAR imagery, were classified individually by a statistical classifier and the neural network, and the results were combined by several voting schemes. In cases where the two classifier outputs disagree, a training sample is rejected and used for the training of a second neural network. In doing so the classification accuracy was significantly increased. However, the requirement of a relatively large training sample set may be considered as a drawback of this method.

Contrary to voting or rejection schemes, in other studies an additional classifier was applied to the individual outputs, to combine the output of the previous classifiers (e.g., Waske and Benediktsson, 2007). In a first processing step, two data sets (i.e., SAR and multispectral) were classified by individual SVM, providing distances between each pixel and the separating decision function (i.e., hyperplane). Afterward an additional SVM was applied on these outputs. The proposed strategy outperforms the fusion with conventional voting schemes as well as a single SVM applied to the full data set or other parametric and nonparametric methods (i.e., maximum likelihood, decision tree, boosted decision tree) as well as two different voting schemes. Besides, the results clearly show the advantage of multisensor imagery in regard to the classification accuracy, which was improved irrespective of the classifier algorithm and training data. In Waske and van der Linden (2008), the approach was further extended by a multilevel component, which enables the inclusion of different segmentation scales of the input imagery. In doing so the classification accuracy was further increased compared to a pixel-wise classification and classifications, which are based on a single segmentation level.

In Udelhoven et al. (2009), the approach was successfully applied to hypertemporal data, consisting 10 day maximum-value NDVI composites from the "Mediterranean Extended Daily One-km Advanced Very High Resolution Radiometer Data Set" (MEDOKADS). Individual SVM classifiers were trained on normalized difference vegetation index (NDVI) time-series and mean annual temperature values of three consecutive years. Afterward the preliminary SVM outputs were fused using an

additional SVM. Compared to other simple voting strategies and a single SVM, which is trained for the same multiannual periods, the classification accuracy and robustness was increased by the proposed method. Moreover, it was demonstrated that considering the interannual variability of the annual vegetation-growth cycle improves the classification accuracy.

Summary

In this entry on decision fusion for classification of multisource remote sensing data, a brief definition on data fusion and an overview of multisource remote sensing applications was given. Examples of how approaches, which are based on decision fusion, can be applied to currently typical remote sensing data sets were presented.

In general, these studies show that the accuracy of remote sensing land cover classifications is increased due to the rapid development of EO systems, resulting in the availability of multitemporal, multisource, or high-dimensional data sets on the one hand and the use of adequate classification approaches on the other. Recent approaches, for example, based on decision fusion, enable the handling of such complex data sets efficiently and in a robust manner. Problems that occur due to limitations of traditional methods can be avoided, and results are consequently more accurate.

A higher number of diverse EO instruments and further enhancement of these systems is expected for the future. Therefore, appropriate data fusion and classification of multisource data is an important ongoing research topic in the field of remote sensing. The methodological development outlined in this entry is an alternative direction to continue.

Bibliography

- Bachmann, C. M., Bettenhausen, M. H., Fusina, R. A., Donato, T. F., Russ, A. L., Burke, J. W., Lamela, G. M., Rhea, W. J., Truitt, B. R., and Porter, J. H., 2003. A credit assignment approach to fusing classifiers of multiseason hyperspectral imagery. *IEEE Transaction on Geoscience and Remote Sensing*, **41**, 2488–2499.
- Benediktsson, J. A., and Kanellopoulos, I., 1999. Classification of multisource and hyperspectral data based on decision fusion. *IEEE Transaction on Geoscience and Remote Sensing*, **37**, 1367–1377.
- Benediktsson, J. A., and Swain, P. H., 1992. Consensus theoretic classification methods. *IEEE Transaction on Systems, Man and Cybernetic*, **22**, 688–704.
- Benediktsson, J. A., Swain, P. H., and Ersoy, O. K., 1990. Neural network approaches versus statistical methods in classification of multisource remote-sensing data. *IEEE Transaction on Geoscience and Remote Sensing*, **28**, 540–552.
- Benediktsson, J. A., Sveinsson, J. R., and Swain, P. H., 1997. Hybrid consensus theoretic classification. *IEEE Transaction on Geoscience and Remote Sensing*, **35**, 833–843.
- Benediktsson, J. A., Pesaresi, M., and Arnason, K., 2003. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transaction on Geoscience and Remote Sensing*, **41**, 1940–1949.
- Blaes, X., Vanhalle, L., and Defourny, P., 2005. Efficiency of crop identification based on optical and SAR image time series. *IEEE Transaction on Geoscience and Remote Sensing*, **96**, 352–365.

- Briem, G. J., Benediktsson, J. A., and Sveinsson, J. R., 2002. Multiple classifiers applied to multisource remote sensing data. *IEEE Transaction on Geoscience and Remote Sensing*, **40**, 2291–2299.
- Brisco, B., and Brown, R. J., 1995. Multidate SAR/TM synergism for crop classification in western Canada. *Photogrammetric Engineering and Remote Sensing*, **61**, 1009–1014.
- Bruzzone, L., and Carlin, L., 2006. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Transaction on Geoscience and Remote Sensing*, **44**, 2587–2600.
- Bruzzone, L., Chi, M., and Marconcini, M., 2006. A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Transaction on Geoscience and Remote Sensing*, **44**, 3363–3373.
- Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Rojo-Alvarez, J. L., and Martinez-Ramon, M., 2008. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Transaction on Geoscience and Remote Sensing*, **46**, 1822–1835.
- Cheung-Wai Chan, J., and Paelinckx, D., 2008. Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotone mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, **112**, 2999–3011.
- Chi, M., and Bruzzone, L., 2007. Semisupervised classification of hyperspectral images by SVMs optimized in the primal. *IEEE Transaction on Geoscience and Remote Sensing*, **45**, 1870–1880.
- Chust, G., Ducrot, D., and Pretus, J. L., 2004. Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape. *International Journal of Remote Sensing*, **25**, 3513–3528.
- Dalla Mura, M., Benediktsson, J. A., Waske, B., and Bruzzone, L., 2010. Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *International Journal of Remote Sensing*, **31**, 5975–5991.
- Dalponte, M., Bruzzone, L., and Gianelle, D., 2008. Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transaction on Geoscience and Remote Sensing*, **46**, 1416–1427.
- De Wit, A. J. W., and Clevers, J. G. P. W., 2004. Efficiency and accuracy of per-field classification for operational crop mapping. *International Journal of Remote Sensing*, **68**(11), 1155–1116.
- Duda, R. O., Hart, P. E., and Stork, D. G., 2001. *Pattern Classification*, 2nd edn. New York: Wiley-Interscience.
- Fauvel, M., Chanussot, J., and Benediktsson, J. A., 2006. Decision fusion for the classification of urban remote sensing images. *IEEE Transaction on Geoscience and Remote Sensing*, **44**, 2828–2838.
- Fauvel, M., Benediktsson, J. A., Chanussot, J., and Sveinsson, J. R., 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, **46**, 3804–3814.
- Fitzgerald, R. W., and Lees, B. G., 1994. Assessing the classification accuracy of multisource remote-sensing data. *Remote Sensing of Environment*, **47**, 362–368.
- Foody, G. M., and Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transaction on Geoscience and Remote Sensing*, **42**, 1335–1343.
- Friedl, M. A., and Brodley, C. E., 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, **61**, 399–409.
- Hall, D. L., and Llinas, J., 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE*, **85**, 6–23.
- Huang, H., Legarsky, H., and Othman, M., 2007. Land-cover classification using Radarsat and Landsat imagery for St. Louis, Missouri. *Photogrammetric Engineering and Remote Sensing*, **73**, 37–43.
- Jain, K., Duin, R. P. W., and Mao, J., 2000. Statistical pattern recognition: a review. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **22**, 4–37.
- Jeon, B., Landgrebe, and D. A., 1999. Decision fusion approach for multitemporal classification. *IEEE Transactions on Geoscience and Remote Sensing*, **37**(3), 1227–1233.
- Koetz, B., Morsdorf, F., van der Linden, S., Curt, T., and Allgöwer, B., 2008. Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. *Forest Ecology and Management*, **256**, 263–271.
- Kuncheva, L. I., 2004. *Combining Pattern Classifiers. Methods and Algorithms*. Hoboken, NJ: Wiley.
- Landgrebe, D., 2003. *Signal Theory Methods in Multispectral Remote Sensing*. New Jersey: Wiley.
- Michelson, D. B., Liljeberg, B. M., and Pilesjö, P., 2000. Comparison of algorithms for classifying Swedish landcover using LANDSAT TM and ERS-1 SAR data. *Remote Sensing of Environment*, **71**, 1–15.
- Pedergnana, M., Marpu, P. R., Dalla Mura, M., Benediktsson, J. A., and Bruzzone, L., 2012. Classification of remote sensing optical and LiDAR data using extended attribute profiles. *IEEE Journal of Selected Topics in Signal Processing*, **6**, 856–865.
- Pohl, C., and van Genderen, J. L., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. *International Journal of Remote Sensing*, **19**, 823–854.
- Polikar, R., 2006. Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, **6**, 21–45.
- Richards, J. A., 2005. Analysis of remotely sensed data: the formative decades and the future. *IEEE Transaction on Geoscience and Remote Sensing*, **43**, 422–432.
- Richards, J. A., and Jia, X., 2006. *Remote Sensing Digital Image Analysis: An Introduction*, 4th edn. New York: Springer.
- Roscher, R., Waske, B., and Förstner, W., 2012. Incremental import vector machines for classifying hyperspectral data. *IEEE Transaction on Geoscience and Remote Sensing*, **50**, 1–11.
- Serpico, S. B., and Roli, F., 1995. Classification of multisensor remote-sensing images by structured neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, **33**(3), 562–578.
- Song, M., Civco, D. L., and Hurd, J. D., 2005. A competitive pixel-object approach for land cover classification. *International Journal of Remote Sensing*, **26**, 4982–4997.
- Steele, B. M., 2000. Combining multiple classifiers: an application using spatial and remotely sensed information for land cover type mapping. *Remote Sensing of Environment*, **74**, 545–556.
- van Genderen, J. L., and Pohl, C., 1994. Image fusion: issues, techniques and applications. Intelligent image fusion. In van Genderen, J. L., and Cappellini, V. (eds.), *Proceedings EARSeL Workshop*, September 11, 1994, Strasbourg, France. Enschede: ITC, pp. 18–26.
- Wald, L., 1999. Some terms of reference in data fusion. *IEEE Transaction on Geoscience and Remote Sensing*, **37**, 1190–1193.
- Waske, B., and Benediktsson, J. A., 2014. Pattern recognition and classification. In Njoku, E. G. (ed.) *Encyclopedia of Remote Sensing*. New York: Springer.
- Waske, B., and Benediktsson, J. A., 2007. Fusion of support vector machines for classification of multisensor data. *IEEE Transaction on Geoscience and Remote Sensing*, **45**, 3858–3866.
- Waske, B., and van der Linden, S., 2008. Classifying multilevel imagery from SAR and optical sensors by decision fusion. *IEEE Transaction on Geoscience and Remote Sensing*, **46**, 1457–1466.
- Waske, B., van der Linden, S., Benediktsson, J. A., Rabe, A., and Hostert, P., 2010. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 2880–2889.
- Wright, C., and Gallant, A., 2007. Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data. *Remote Sensing of Environment*, **107**, 582–605.

E

EARTH RADIATION BUDGET, TOP-OF-ATMOSPHERE RADIATION

Bing Lin
NASA Langley Research Center, MS 420, Hampton,
VA, USA

Synonyms

Extraterrestrial radiation; Longwave radiation; Shortwave radiation; Solar radiation; Terrestrial radiation; Thermal radiation

Definition

1. Extraterrestrial radiation: the radiative energy that radiates into the Earth's climate system from the space.
2. Terrestrial radiation: the radiative energy emitted by the Earth's climate system.

Introduction

The radiation at top of the atmosphere (TOA) is the only energy exchange process between the Earth and space for the Earth's climate system. The incoming extraterrestrial radiation, dominantly from the sun, is the driving forcing for the general circulation of the climate system. In response to this solar radiative forcing, the climate system releases heat to the space through terrestrial thermal emission. For the Earth, the incoming solar radiation is generally balanced by the outgoing thermal radiation at TOA at the annual and longer timescales to maintain the climate system in a quasi-equilibrium state. Some anthropogenic processes, such as the continuous increase of the CO₂ concentration within the atmosphere caused by the increased fossil fuel burning, inject extra thermal radiative forcing into the climate system, which breaks the basic radiation balance and forces the climate to change. Since the coupling of solar and thermal radiative energy

processes with the Earth's hydrological processes such as precipitation and evaporation builds a fundamental aspect of the global water and energy cycle of the climate system, variations in radiative fields would cause corresponding changes in global precipitation patterns and freshwater distributions and have a strong social-economical impact. The observations of TOA radiation, thus, become a critical part of the remote sensing of the Earth's climate system.

The top-of-the-atmosphere radiation

Satellite direct measurements of TOA radiation (Barkstrom et al., 1989; Wielicki et al., 1996) focus on broadband spectra of the total (0.2 ~ 100 μm), shortwave (0.2 ~ 5 μm), and longwave (5 ~ 100 μm) radiation, which cover more than 99.5 % of the Earth's radiant energy. Absolute calibration and long-term stability are the core elements emphasized by satellite TOA radiation-measuring missions. Both wide and narrow field of view (FOV) techniques have been used in satellite TOA radiation measurements for decades (Barkstrom et al., 1989; Wielicki et al., 1996). Although wide FOV observations can obtain the information of global net radiation, detailed radiative transfer processes for regional and small-scale phenomena such as clouds and aerosols cannot be measured. The narrow FOV measurements, on the other hand, have reasonable spatial resolution in monitoring all large, regional, and small-scale physical processes. But to obtain the full radiant energy of these physical processes, a series of complicated satellite retrievals is required for narrow FOV radiance measurements. Each instantaneous satellite observation of the Earth's radiation field from narrow FOV broadband radiation instruments can only measure the radiance at a particular viewing angle of the targeted scene, and each scene type has its distinct three-dimensional (3-D) radiation features. In order to retrieve the radiative flux of the targeted scene from the radiance

measurement at the particular viewing angle, a 3-D angular dependent model (ADM) of radiance field for the scene type has to be used. The ADM is built from the statistics of large amounts of observational multi-angle radiance data. Thus, the spatial resolution of radiation measurements should be both large enough to satisfy the 3-D radiation statistics of individual scene types and fine enough to determine the radiative characteristics of major climate processes. Satellite radiation sensors, such as those from the Earth Radiation Budget Experiment (ERBE) (Barkstrom et al., 1989) and Clouds and the Earth's Radiant Energy System (CERES) (Wielicki et al., 1996), normally have a spatial resolution from 10 to 40 km and generally not only measure the fluxes of total, shortwave, and longwave radiation at TOA but also differentiate the radiation from individual Earth's scene types, such as clouds, aerosols, clear skies, deserts, grasses, forests, and ice sheets. Temporal resolutions of satellite radiation measurements depend on both satellite orbits and scan patterns of radiation sensors. Precession orbits can provide radiative flux measurements of broad solar incident angles for each place and scene type but usually have low repeating sampling rate. Polar orbital sensors, on the other hand, have high repeating rates of measurements for similar weather and climate conditions, but their measurements are only available for limited local times.

Another way in satellite estimations of TOA radiant energy is to observe the radiative properties of all critical agents and components within the atmosphere and at surfaces through satellite narrowband radiation measurements such as those from passive and active visible, infrared, and microwave remote sensing techniques and then to calculate the radiant energy based on these radiative properties using radiative transfer theories. The spatial and temporal resolutions of these radiative flux estimates, thus, depend on their corresponding narrowband satellite measurements, which have a wide range of variations. As the methods of satellite direct broadband radiation measurements, this indirect measurement method has also been used for decades (Rossow and Schiffer, 1999) and is a major part of the satellite remote sensing of global radiation budgets.

Satellite radiative energy measurements track all important physical processes within the Earth's climate system and are currently one of the keys in monitoring and understanding of the climate system. Climate data records from the satellite TOA radiation measurements capture global, regional, and small-scale signatures of the climate variability and change and exhibit distinct features of the tropics and polar regions in response to global warming trends. These TOA radiation measurements, especially those for cloudy and clear skies, also provide strong energy constraints for global climate models (GCMs) and lead advanced GCM cloud parameterizations. With the continuous, long-term radiation observations from CERES and other satellite radiation missions, an improved projection of future climate could be obtained.

Summary

Satellite measurements of the Earth's radiation budget have significant influences on the understanding of aerosols, cloud feedbacks, and atmospheric dynamics as well as other major climate physical processes and are critical for validations and improvements of climate models. Continuous, long-term radiation observations from satellite missions would lead to accurate predictions of the future climate.

Bibliography

- Barkstrom, B., et al., 1989. The Earth Radiation Budget Experiment (ERBE) archival and April 1985 results. *Bulletin of the American Meteorological Society*, **70**, 1254–1262.
- Rossow, W. B., and Schiffer, R. A., 1999. Advances in understanding clouds from ISCCP. *Bulletin of the American Meteorological Society*, **80**, 2261–2287.
- Wielicki, B. A., Barkstrom, B., Harrison, E. F., Lee, R., Smith, G., and Cooper, J., 1996. Clouds and the Earth's Radiant Energy System (CERES): An Earth observing system experiment. *Bulletin of the American Meteorological Society*, **77**, 853–868.

Cross-references

[Aerosols](#)
[Atmospheric General Circulation Models](#)
[Calibration and Validation](#)
[Climate Data Records](#)
[Climate Monitoring and Prediction](#)
[Cloud Properties](#)
[Optical/Infrared, Radiative Transfer](#)
[Radiation \(Natural\) Within the Earth's Environment](#)
[Surface Radiative Fluxes](#)
[Water and Energy Cycles](#)

EARTH SYSTEM MODELS

Andrea Donnellan
 Science Division, Jet Propulsion Laboratory, California
 Institute of Technology, Pasadena, CA, USA

Definition

Earth system. Various parts of the Earth that interact with each other.

Models. A representation, either physically, graphically, or using a computer, of a scientific process or phenomenon developed for providing understanding of the real process or phenomenon.

Earth system model. A representation through graphics, computer software, or physically of interacting parts of the Earth developed to provide improved understanding.

Introduction

The solid Earth includes the core, mantle, lithosphere and crust, and the cryosphere consisting of ice sheets and glaciers. The core is liquid with a solid middle and generates our Earth's magnetic field. The mantle is made of hot rock

that convects, driving plate tectonics in the lithosphere and crust. Earthquakes and volcanoes occur as a result of plate tectonics. The ice masses that make up the cryosphere interact with the solid Earth, oceans, and atmosphere.

Satellite observations are used to probe characteristics of the solid Earth from the outer liquid core to the Earth's surface.

Core

The Earth's core is made up of a solid inner core and a liquid outer core. The outer core is composed of a liquid metal alloy and forms a layer 2,260 km thick. The electrically conducting liquid metal alloy rotates and convects, generating the Earth's magnetic field. The motions within the liquid outer core vary, resulting in both long-term and short-term changes in the magnetic field of the Earth. The Earth's magnetic dipole or north and south poles are generated from the outer core and change slowly, resulting in magnetic reversals that occur on timescales of millions of years. Shorter-term variations are superposed on the long-term variations.

Measurements of the magnetic field provide a way of probing the properties of the flow in the outer core. It is necessary to separate out the magnetic field from the magnetized rocks in the Earth's crust, electrical currents from the ocean, atmosphere, and ionosphere, as well as external fields. Measurements that have been carried out over the last 150 years indicate that the Earth's magnetic field has decayed by 10 %, consistent with an ongoing reversal of the magnetic poles. The bulk of the decay is occurring at the South Atlantic Ocean. Remote sensing can detect magnetic variations on the order of 1 year, which provides information on how the fluids move within the outer core. Remote sensing measurements can also further separate the internal and external processes that generate the Earth's magnetic field. Understanding the motion of the fluids in the outer core will result in a better understanding of the motions of the oceans and atmospheres and their circulation by allowing that contribution to be separated out from the observed total angular momentum budget of the Earth.

Mantle

The mantle is a layer about 2,900 km thick between the core and the crust. The mantle is composed primarily of crystalline silicate rock, which is hot and also convects. Seismic data and gravity data collected by remote sensing indicate that there is a discontinuity in material at about 400–670 km, which marks the boundary between the upper and lower mantle. The seismic waves are deflected by the discontinuity. Scientists predict the gravitational field of the Earth based on what they know about the distribution of densities, and they then compare their prediction to ground- and/or space-based measurements of the gravitational field of the Earth. Gravity anomalies are the differences between the observed and predicted gravity field. The convection of the mantle results in varying mass

distributions throughout the mantle. Gravity anomalies have been used to show that the viscosity of the mantle increases dramatically at the boundary between the upper and lower mantle, suggesting that convection occurs in the upper mantle separately from the lower mantle.

The asthenosphere makes up the upper part of the mantle and is about 200 km thick. The rocks in the asthenosphere flow readily and are typically made up of olivine. The lithosphere, which is marked at the bottom by the transition from ductile to brittle olivine, makes up the very top of the mantle and the lower part of the Earth's crust and is made up of rigid rocks. Convection of the mantle drives plate tectonics and molten rock from the mantle, comes to the surface by way of volcanoes, and spreads ridges between the tectonic plates.

Lithosphere and crust

The rigid lithosphere is typically 50–200 km thick under the oceans and is up to 200 km thick under the continents. The Earth's crust is found in the top 30–50 km of the lithosphere over continents. The chemical composition of the rocks changes in the crust, and the base is marked by the Moho discontinuity, determined from seismology.

Plate tectonics

The mantle convection that drives plate tectonics creates convergent zones, where one plate subducts under another plate, transform, and extensional zones. Transform plate boundaries are typically found in the ocean but can also occur on land, such as in California. Extension is typically found at spreading ridges in the ocean. Hot spots, which are upwellings of hot material from deep in the mantle, can also cause spreading in the crust. The Basin and Range Province in the Western United States is an example of spreading induced by a hot spot. Volcanoes with different characteristics can form over hot spots, spreading centers, and subduction zones.

Seismology has been used extensively to understand earthquakes as well as characteristics of the solid Earth and of volcanoes. The advent of remote sensing has allowed for the measurement of surface motions to millimeter precision. This makes it possible to measure plate tectonics, comparing the rates to those determined through other geologic techniques. Measurement of surface deformation yields information about the material properties such as strengths and rigidity of the lithosphere and mantle and provides insight into the complete strain accumulation and release related to earthquakes and volcanoes.

Earthquakes

Earthquakes occur in the brittle part of the crust and mantle. They occur when stress increases, typically from motion due to the tectonic plates, and exceeds a threshold, at which point the crust breaks. The resulting shaking is an earthquake, which can be measured through the use of seismometers. However, the earthquake makes up only a small part of the entire earthquake cycle of strain

accumulation and release. Earthquakes typically occur along faults that have broken many times in the past. They are associated with convergent, transform, and extensional plate tectonics. There are three major types of faults, which are thrust, strike-slip, and normal faults.

Thrust faults and thrust earthquakes occur where the crust is shortening or converging. In these types of earthquakes, one block slides up over the bottom block along a fault oriented diagonally from depth toward the surface. Thrust earthquakes typically occur at subduction zones but can also occur where there is mantle downwelling or plate collision underneath the crust. Subduction zones are found along the margins of the Pacific plate along the coasts of South America, the Pacific Northwest of the United States, Canada, and Alaska, and underneath Japan. A prime example of a collision zone is the Himalayas. The thrust earthquakes in Southern California are associated with a downwelling underneath the Transverse Ranges.

Strike-slip earthquakes result when one side of a fault slips horizontally in either direction on both sides of a fault. Strike-slip faults occur along transform boundaries that offset spreading centers along divergent plate boundaries. The San Andreas Fault in California is a major strike-slip fault and makes up the primary boundary between the Pacific and North American Plates. The actual plate boundary is about 100 km wide, however.

Normal faults occur in regions where extension is occurring. A normal fault is also oriented diagonally from the upper part to the lower part at depth, but in this case, the upper block moves down and away relative to the lower block along the fault. The Basin and Range Province of the Western United States produces many normal faults. As the crust stretches, blocks in the valley, known as grabens, move down. The mountains move upward and are referred to as horsts.

Various techniques can be used to measure the strain associated with the earthquake cycle. Ground-based observations provided measurement of strain. The Global Positioning System (GPS) is now used to measure surface deformation at mm/year accuracy around the globe. The measurements provide time continuous precise measurements at distinct points. Another technique, called interferometric synthetic aperture radar (InSAR), relies on radar waves transmitted from a satellite and reflected back to the satellite from the Earth's surface. When the satellite exactly repeats its track, then differences in the radar return are due to surface motions, as well as error sources. The two observations are coupled together and form an interferogram where the radar waves interfere due to the surface motions. InSAR provides a precise map of deformation of the Earth's surface to/from the satellite.

Faults typically creep at depth, where the crust or mantle is hotter, and are locked at the surface except for when they break in earthquakes. The brittle part of the upper crust behaves elastically and moves in a predictable manner when strained. Observations of surface deformation yield information about the entire earthquake cycle, the

cycle of strain accumulation and release, and provide information about fault activity and slip rates.

Recent work using remote sensing has demonstrated that silent slip events that mimic earthquake slip without the shaking are fairly common. Additionally a substantial amount of continued motion, referred to as postseismic response, can occur following earthquakes. The postseismic motions can result from relaxation of the more ductile lower crust or continued slip on the main fault plane. Models that incorporate seismic and crustal deformation are providing a system level understanding of earthquake processes and are improving assessments of earthquake hazards.

Volcanoes

Volcanoes are ruptures at the Earth's surface that allow magma and gases to escape from the Earth's interior. There are three primary types of volcanoes associated with divergent and convergent plate boundaries and hot spots. At spreading centers associated with divergent plate boundaries, the crust is typically thin, and the magma is of low viscosity and basaltic, coming from melted mantle. Subduction zone volcanoes associated with divergent plate boundaries tend to be explosive as the high-silicate-content subducting crust is heated and rises. A rising plume from a mantle hot spot also creates volcanoes. In cases such as Hawaii, chains of volcanoes form as the tectonic plate moves over the relatively fixed hot spot.

Magma moving beneath the surface of the Earth causes the surface of the Earth to deform and change temperature. GPS and interferometric radar (InSAR) are used to measure surface deformation, while short-wavelength infrared data are used to measure thermal characteristics of the surface. These observations are used to infer the location and migration of magma at depth, which are then used to determine the activity of a volcano and likelihood of eruption.

Cryosphere

Ice masses covering the Earth make up the cryosphere. This includes ice sheets, glaciers, and sea ice. The cryosphere has shown tremendous variability in the geologic record. The components of the cryosphere are sensitive to inputs and thermal exchanges from the atmosphere, ocean, and solid Earth. Volcanoes, hot spots, and subsurface topography, and geology at the base can control behavior of ice masses that overlie them. Ocean sea level and temperature affect the margins of ice sheets, which can lead to changes in the interiors. Changes in atmospheric temperature and precipitation can also modify ice sheet behavior. Remote sensing can be used to understand long-term ice sheet and glacier records as well as present-day ice dynamics.

Glacio-isostatic adjustment

Ice sheets depress the crust and mantle that underlie them. As ice sheets thicken, the surface of the rock is depressed downward, flexing the crust beyond the extent of the ice

sheet. When ice sheets thin, the opposite occurs and the crust and mantle rebound. However, the mantle is viscous and as a result has a delayed response for complete rebound. Knowing the mantle viscosity allows inferences to be made about past ice sheet history. Conversely, understanding ice sheet history can provide insight into mantle viscosity.

GPS measurements provide the rates and distribution of present-day glacio-isostatic adjustment. Time-varying gravity measurements from space indicate locations of mass changes, which can be related to glacio-isostatic adjustment. These measurements are used in Earth system models to understand past ice sheet history and crust and mantle properties.

Ice sheet dynamics

Ice sheets are quite variable on short timescales as well as on long timescales. The mass balance of an ice sheet refers to whether the ice sheet is thickening or thinning (losing or gaining mass). Understanding the mass balance of ice sheets requires knowledge of the thickness changes of the ice sheet, but also the flow rates of the ice sheet. Repeat altimetry, such as laser altimetry from LIDAR (Light Detection and Ranging), informs scientists regarding height changes of the ice sheet, which can be translated to thickness changes. Adding in flow measured from GPS or InSAR allows for a complete measurement of the volume changes of the ice sheet. Flow can also be inferred from tracking features in visible images. It is also possible to observe where the changes are occurring and whether those changes are migrating with time. For example, the Greenland ice sheet measurements indicate that the thinning started near the coast and is migrating inland, suggesting that ocean changes are a bigger driver than atmospheric changes. Earth system models take into account the basal characteristics of the ice sheet, flow and thickness parameters, as well as ocean and atmospheric coupling.

Ice shelves floating on the margins of ice sheets can be quite unstable. Radar and other imagery can be used to observe breakup of ice shelves as it occurs. The breakup of ice shelves also needs to be factored into flow models of the adjacent ice sheet.

Sea ice

Changes in sea ice cover provide indicators of how the climate is changing and also impacts shipping lanes. Radar and other imagery can be used to track sea ice flow, coverage characteristics, and floe distributions. Altimetry, such as from LIDAR, can be used to estimate the sea ice freeboard, or the amount of sea ice floating above the surface. Passive microwave remote sensing is used to measure the thermal, dielectric, and geophysical properties of snow on the sea ice. The measured parameters coupled with models show the annual evolution of the Arctic sea ice, atmosphere, and ocean system.

Conclusions

The solid Earth functions as a system, ranging from the core through the mantle to the crust and cryosphere. The geodynamo operating within the core provides the Earth's magnetic field, which varies due to flow within the core. The mantle interacts with and is a key component of plate tectonic, earthquake, and volcanic systems, and earthquakes and volcanoes result from plate tectonics. The cryosphere also interacts with the crust and the mantle depressing the solid Earth surface due to the mass of ice sheets. Tectonic plate motions can in turn affect ice sheets. It is important to model the Earth as a system in order to separate out different effects and to understand the individual components of the system.

Remote sensing offers a way of probing various components of the solid Earth, providing key parameters to Earth system models. The mechanisms that drive various solid Earth processes primarily occur underneath the surface and as such are not directly observable. Furthermore, remote sensing provides a regional view that is more difficult to obtain with in situ measurements.

Acknowledgments

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Bamber, J. L., and Rivera, A., 2007. A review of remote sensing methods for glacier mass balance determination. *Global and Planetary Change*, **59**(1–4), 138–148.
- Braun, A., Kuob, C.-Y., Shum, C. K., Wu, P., van der Wal, W., and Fotopoulos, G., 2008. Glacial isostatic adjustment at the Laurentide ice sheet margin: models and observations in the Great Lakes region. *Journal of Geodynamics*, **46**(3–5), 165–173.
- Bürgmann, R., Rosen, P. A., and Fielding, E. J., 2000. Synthetic aperture radar interferometry to measure earth's surface topography and its deformation. *Annual Review of Earth and Planetary Sciences*, **28**, 169–209.
- Board on Atmospheric Sciences and Climate, 2008. *Earth Observations from Space: The First 50 Years of Scientific Achievements*.
- Delouis, B., Giardini, D., Lundgren, P., and Salichon, J., 2002. Joint inversion of InSAR, GPS, teleseismic, and strong-motion data for the spatial and temporal distribution of earthquake slip: application to the 1999 Izmit Mainshock. *Bulletin of the Seismological Society of America*, **92**(1), 278–299.
- Donnellan, A., Parker, J. W., and Peltzer, G., 2002. Combined GPS and InSAR models of postseismic deformation from the Northridge earthquake. *Pure and Applied Geophysics*, **159**, 10.
- Glaze, L., Francis, P. W., and Rothery, D. A., 1989. Measuring thermal budgets of active volcanoes by satellite remote sensing. *Nature*, **338**, 144–146.
- Hager, B. H., 1984. Subducted slabs and the geoid: constraints on mantle rheology and flow. *Journal of Geophysical Research*, **89**, 6003–6016.
- <http://www.es.ucsc.edu/~glatz/geodynamo.html>
- http://www.esa.int/esaLP/ESA3QZJE43D_LPswarm_0.html
- Jeanloz, R., 1990. The nature of the earth's core. *Annual Review of Earth and Planetary Sciences*, **18**, 357–386.

- Korsnes, R., 1993. Quantitative analysis of sea ice remote sensing imagery. *International Journal of Remote Sensing*, **14**(2), 295–311.
- Langlois, A., and Barber, D. G., 2007. Passive microwave remote sensing of seasonal snow-covered sea ice. *Progress in Physical Geography*, **31**, 539.
- Rogers, G., and Dragert, H., 2003. Episodic tremor and slip on the cascadia subduction zone: the chatter of silent slip. *Science*, **300**(5627), 1942–1943.
- Segall, P., Desmarais, E. K., Shelly, D., Miklius, A., and Cervelli, P., 2006. Earthquakes triggered by silent slip events on Kilauea volcano, Hawaii. *Nature*, **442**, 71–74.
- Solomon, S. C., 2002. *Living on a Restless Planet. Solid Earth Science Working Group Report*. National Aeronautics and Space Administration. JPL 400–1040.
- Stevenson, D. J., 1981. Models of the earth's core. *Science*, **214**, 4521.
- Tralli, D. M., Blom, R. G., Zlotnicki, V., Donnellan, A., and Evans, D. L., 2005. Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *ISPRS Journal of Photogrammetry and Remote Sensing*, **59**(4), 185–198.
- Zebker, H. A., Falk, A., and Jonsson, S., 2000. Remote sensing of volcano surface and internal processes using radar interferometry. *Geophysical Monograph*, **116**, 179–205.

Cross-references

[Atmospheric General Circulation Models](#)
[Calibration and Validation](#)
[Calibration, Microwave Radiometers](#)
[Calibration, Optical/Infrared Passive Sensors](#)
[Calibration, Scatterometers](#)
[Calibration, Synthetic Aperture Radars](#)
[Climate Data Records](#)
[Climate Monitoring and Prediction](#)
[Cryosphere and Polar Region Observing System](#)
[Cryosphere, Climate Change Effects](#)
[Cryosphere, Climate Change Feedbacks](#)
[Cryosphere, Measurements and Applications](#)
[Data Access](#)
[Data Archival and Distribution](#)
[Data Archives and Repositories](#)
[Data Assimilation](#)
[Data Processing, SAR Sensors](#)
[Decision Fusion, Classification of Multisource Data](#)
[Emerging Technologies, LIDAR](#)
[Emerging Technologies, Radar](#)
[Emerging Technologies, Sensor Web](#)
[Geodesy](#)
[Geomorphology](#)
[Geophysical Retrieval, Overview](#)
[Geophysical Retrieval, Inverse Problems in Remote Sensing](#)
[Global Earth Observation System of Systems \(GEOSS\)](#)
[Global Land Observing System](#)
[Ice Sheets and Ice Volume](#)
[Magnetic Field](#)
[Ocean Modeling and Data Assimilation](#)
[Remote Sensing, Physics and Techniques](#)
[Sea Ice Albedo](#)
[Sea Ice Concentration and Extent](#)
[Sea Level Rise](#)
[Solid Earth Mass Transport](#)
[Subsidence](#)

ELECTROMAGNETIC THEORY AND WAVE PROPAGATION

Yang Du

Zhejiang University, Hangzhou, People's Republic of China

Definition

Electromagnetic theory. Theory based on the Maxwell equations to study the properties and dynamics of electromagnetic fields.

Wave propagation. The process of transporting electromagnetic energy.

Maxwell equations

The Maxwell equations lay the foundation of all classical electromagnetic phenomena. They describe how time-varying electric fields give rise to magnetic fields and vice versa. The Maxwell equations are a set of four equations

$$\nabla \cdot \mathbf{D} = \rho \text{ (Coulomb's law)}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \text{ (Faraday's law)} \quad (1)$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}$$

(Ampere's law with displacement current addition)

$$\nabla \cdot \mathbf{B} = 0 \text{ (Absence of free magnetic poles)}$$

The six quantities in terms of which the electromagnetic field equations are expressed are as follows:

- $\mathbf{D}$ = Electric flux density or electric displacement (coulombs per square meter)
- $\mathbf{E}$ = Electric intensity (volts per meter)
- $\mathbf{B}$ = Magnetic flux density (Weber's per square meter)
- $\mathbf{H}$ = Magnetic intensity (amperes per meter)
- $\mathbf{J}$ = Electric current density (amperes per square meter)
- ρ = Electric charge density (coulombs per cubic meter)

In vacuum $\mathbf{D} = \epsilon_0 \mathbf{E}$ and $\mathbf{B} = \mu_0 \mathbf{H}$, where $\epsilon_0 \approx 8.85 \times 10^{-12}$ farad/m is the permittivity, and $\mu_0 = 4\pi \times 10^{-7}$ Henry/m is the permeability. The continuity equation for charge density and current density can be derived from the above Maxwell equations

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (2)$$

When electromagnetic wave propagates in some medium other than vacuum, since the characteristic wavelength is very large compared to the atoms of which the medium is composed, the detailed behavior of the fields over atomic distance can be neglected. What are relevant are the quantities averaged over the atomic scale, including the macroscopic fields and macroscopic

sources. The constitutive relations, which relate $\mathbf{D}$ and $\mathbf{B}$ to $\mathbf{E}$ and $\mathbf{H}$, may take forms much more complicated than those in vacuum (see [Media, Electromagnetic Characteristics](#)).

Boundary conditions

Maxwell equations in the form of [Equation 1](#) concern ordinary points of space in whose neighborhood the physical properties of the medium vary continuously. However, at boundary of two different media, the permittivity, permeability, and conductivity can change sharply. The discontinuous change on a macroscopic scale will lead to corresponding discontinuity in the field vectors. The discontinuity equations are collectively called boundary conditions (Jackson, 1998; Kong, 2005).

The boundary conditions for the normal components of the electric flux density $\mathbf{D}$ and magnetic flux density $\mathbf{B}$ on either side of the boundary surface are

$$\begin{aligned} (\mathbf{D}_1 - \mathbf{D}_2) \cdot \hat{n} &= \sigma \\ (\mathbf{B}_1 - \mathbf{B}_2) \cdot \hat{n} &= 0 \end{aligned} \quad (3)$$

where the normal $\hat{n}$ is pointing from medium 1 into medium 2, σ is an idealized surface current density that is related to the charge density ρ singular at the interface by $\int_{\Delta V} \rho dv' = \sigma \Delta s$, and ΔV is an infinitesimal

pillbox of area Δs and height Δh , with half of its volume in each medium and its top and bottom parallel to the surface. [Equation 3](#) states that at any point on the boundary the normal component of $\mathbf{B}$ is continuous, while that of $\mathbf{D}$ is discontinuous with the discontinuity equal to the surface charge density.

The boundary conditions for the tangential components of $\mathbf{E}$ and $\mathbf{H}$ on either side of the boundary surface are

$$\begin{aligned} \hat{n} \times (\mathbf{E}_2 - \mathbf{E}_1) &= 0 \\ \hat{n} \times (\mathbf{H}_2 - \mathbf{H}_1) &= \mathbf{J}_s \end{aligned} \quad (4)$$

where $\mathbf{J}_s$ is an idealized surface current density flowing on the boundary surface which satisfies the condition $\int_{\Delta S} (\frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}) \cdot \hat{t} ds = \hat{t} \cdot \mathbf{J}_s \Delta l$ and ΔS is an open surface

spanning an infinitesimal contour of line Δl parallel to the surface, with half of its area in each medium and its normal $\hat{t}$ tangent to the surface. [Equation 4](#) states that at any point on the boundary the tangential component of $\mathbf{E}$ is continuous, while that of $\mathbf{H}$ is discontinuous by an amount of identical magnitude with direction parallel to $\mathbf{J}_s \times \hat{n}$.

Scalar and vector potentials

Potentials are useful auxiliary functions in analysis of an electromagnetic field. This is because coupling between the electric and magnetic fields in the Maxwell equation makes direct solution difficult, except in simple situations. Potentials, however, if introduced properly, can lead to decoupled equations of their own while identically

satisfying some of the Maxwell equations. The potentials commonly introduced are the scalar potential ϕ and vector potential $\mathbf{A}$ (Jackson, 1998; Kong, 2005).

Since the magnetic flux density $\mathbf{B}$ is always solenoidal, i.e., $\nabla \cdot \mathbf{B} = 0$, it can be represented as the curl of the vector potential $\mathbf{A}$.

$$\mathbf{B} = \nabla \times \mathbf{A} \quad (5)$$

This representation, together with Faraday's law, implies that the electric intensity $\mathbf{E}$ can be expressed in terms of the scalar and vector potentials as

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} \quad (6)$$

Consider the case where the medium is homogeneous and isotopic, for which the constitutive relations are $\mathbf{D} = \epsilon \mathbf{E}$, $\mathbf{B} = \mu \mathbf{H}$. The arbitrariness of choice of $\mathbf{A}$ inherent in [Equation 5](#) and of ϕ through [Equation 6](#) allows the choice of a set of potentials $(\mathbf{A}, \phi)$ to satisfy the Lorenz condition,

$$\nabla \cdot \mathbf{A} + \mu\epsilon \frac{\partial \phi}{\partial t} = 0 \quad (7)$$

Making use of the remaining two Maxwell equations leads to two inhomogeneous wave equations, uncoupled for ϕ and $\mathbf{A}$ as

$$\begin{aligned} \Delta^2 \phi - \mu\epsilon \frac{\partial^2 \phi}{\partial t^2} &= -\frac{\rho}{\epsilon} \\ \Delta^2 \mathbf{A} - \mu\epsilon \frac{\partial^2 \mathbf{A}}{\partial t^2} &= -\mu \mathbf{J} \end{aligned} \quad (8)$$

These equations, equivalent to the Maxwell equation in all respect in a homogeneous and isotopic medium, can be conveniently analyzed using the Green's function.

Green's function for the wave equation

The underlying structures are identical for the wave equations (Jackson, 1998)

$$\nabla^2 \psi - \mu\epsilon \frac{\partial^2 \psi}{\partial t^2} = -f(\mathbf{r}, t) \quad (9)$$

with $f(\mathbf{r}, t)$ being a known source distribution. After Fourier transform, $\psi(\mathbf{r}, \omega)$ satisfies the inhomogeneous Helmholtz wave equation

$$(\nabla^2 + k^2)\psi(\mathbf{r}, \omega) = -f(\mathbf{r}, \omega). \quad (10)$$

The corresponding Green's function $g(\mathbf{r}, \mathbf{r}')$ satisfies the inhomogeneous equation

$$(\nabla^2 + k^2)g(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}') \quad (11)$$

where $\delta(\mathbf{r})$ is the Dirac delta function. In an unbounded medium, the Green's function is

$$g(\mathbf{r}, \mathbf{r}') = \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \quad (12)$$

where $k = \omega\sqrt{\mu\epsilon}$ is the wave number in the medium and $|\mathbf{r} - \mathbf{r}'|$ is the distance between the field point $\mathbf{r}$ and the source point $\mathbf{r}'$. This form of Green's function represents a diverging spherical wave propagating from the origin.

The above Green's function is the scalar Green's function because it relates scalar field to scalar source. If both field and source are vectors, dyadic Green's function must be used instead which satisfies the inhomogeneous equation (Kong, 2005)

$$\nabla \times \nabla \times \bar{\bar{G}}(\mathbf{r}, \mathbf{r}') - k^2 \bar{\bar{G}}(\mathbf{r}, \mathbf{r}') = \bar{\bar{I}}\delta(\mathbf{r} - \mathbf{r}') \quad (13)$$

where $\bar{\bar{I}}$ is the identity dyad. The dyadic Green's function can be expressed in terms of the scalar Green's function as

$$\bar{\bar{G}}(\mathbf{r}, \mathbf{r}') = \left[\bar{\bar{I}} + \frac{1}{k^2} \nabla \nabla \right] g(\mathbf{r}, \mathbf{r}') \quad (14)$$

Poynting's theorem and conservation of energy

Poynting's theorem is a law of conservation of energy for the electromagnetic field. From the Maxwell equations, considering the fields in a finite volume V , the following equation can be readily established (Jackson, 1998):

$$-\int_V \mathbf{J} \cdot \mathbf{E} dv' = \int_V \left[\nabla \cdot (\mathbf{E} \times \mathbf{H}) + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} \right] \quad (15)$$

Each term of the above equation has physical meaning. The left side of the equation represents the negative of the total work done by the fields on the sources within the volume. Integral of the first term on the right side of the equation represents the energy that flows out of the volume through the boundary surfaces per unit time. Since energy flow is such an important concept, to represent it a vector $\mathbf{S}$ is specifically defined which is called the Poynting vector given by

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}. \quad (16)$$

The last two terms on the right side are collectively associated with the total energy density u (electric and magnetic energy density combined) for a linear medium with negligible dispersion or losses. To be specific, they are the time rate of change of the total energy density, i.e., $\frac{\partial u}{\partial t}$. Integral of the last two terms is then the time rate of change of electromagnetic energy within a certain volume. Equation 15 is a statement of conservation of energy; the diminution of electromagnetic energy stored in V is partly due to electromagnetic energy flows outward across the boundary surfaces and partly due to a conversion of electromagnetic energy into mechanical or heat energy.

For more general case of linear dispersive media, interpretation of the right side of Equation 15 is more involved.

For harmonic electromagnetic fields, it is more convenient to use the complex Poynting vector given by $\mathbf{S} = \frac{1}{2}(\mathbf{E} \times \mathbf{H}^*)$.

Plane wave

Plane waves are the simplest and fundamental electromagnetic waves. All wave functions can be expressed as superpositions of plane waves. In an unbounded uniform isotropic linear media, where $\mathbf{D} = \epsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$, the harmonic plane wave fields are

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \mathbf{E}e^{i\mathbf{k} \cdot \mathbf{r} - i\omega t} \\ \mathbf{H}(\mathbf{r}, t) &= \mathbf{H}e^{i\mathbf{k} \cdot \mathbf{r} - i\omega t}, \end{aligned} \quad (17)$$

which satisfy the homogeneous Helmholtz wave equation. The Maxwell equations demand that both $\mathbf{E}$ and $\mathbf{H}$ are perpendicular to the vector $\mathbf{k}$. If $\mathbf{k}$ is real, then it can be written as $\mathbf{k} = k\hat{\mathbf{k}}$, where $k = \omega\sqrt{\mu\epsilon}$ is the wave number and $\hat{\mathbf{k}}$ is a real unit vector indicating the direction of propagation. If $\mathbf{k}$ is complex, then it is expressed through two real vectors as $\mathbf{k} = \beta + i\alpha$, and the plane wave represented by Equation 17 propagates in the direction of β and attenuates in the direction of α (Harrington, 1961).

If the media are semi-infinite with different properties separated by a plane surface, then at the interface electromagnetic waves experience reflection and refraction. Angle of reflection θ_r equals angle of incidence θ_i , while angle of refraction θ_t is determined by Snell's law, which states that the ratio of sine of angle of refraction to sine of angle of incidence equals the ratio of index of refraction n in the medium of incidence to index of refraction n' in the medium of transmission. The Fresnel reflection coefficient, which specifies the ratio of the reflected electric intensity to the incident electric intensity, depends on how the electric field is directed. If the electric field is perpendicular to plane of incidence, then the Fresnel reflection coefficient is (Jackson, 1998)

$$R_{\perp} = \frac{\cos \theta_i - \frac{\mu}{\mu'} \frac{n'}{n} \cos \theta_t}{\cos \theta_i + \frac{\mu}{\mu'} \frac{n'}{n} \cos \theta_t} \quad (18)$$

where μ and μ' are the permeability in the medium of incidence and of transmission, respectively. If the electric field is parallel to plane of incidence, then the Fresnel reflection coefficient is

$$R_{//} = \frac{\frac{\mu}{\mu'} \frac{n'}{n} \cos \theta_i - \cos \theta_t}{\frac{\mu}{\mu'} \frac{n'}{n} \cos \theta_i + \cos \theta_t}. \quad (19)$$

There is an important aspect associated with this case; when angle of incidence equals Brewster's angle, there is no reflected wave. Brewster's angle can be simply determined by $\theta_B = \tan^{-1}(\frac{\mu}{n})$ when $\mu = \mu'$.

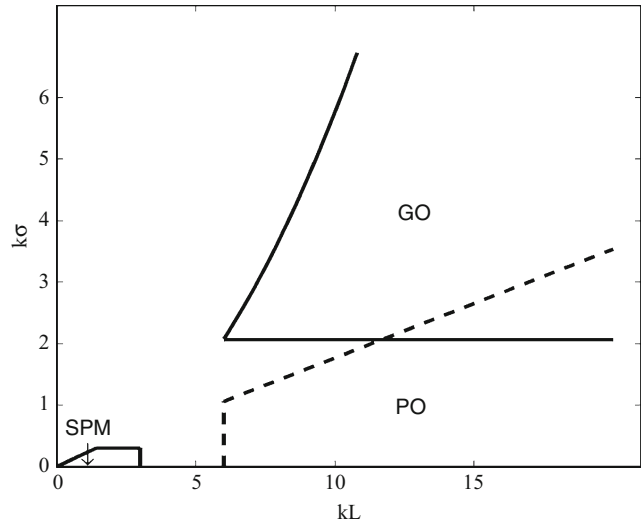
If index of refraction in the medium of incidence is larger than that in the medium of refraction, the phenomenon called total internal reflection can occur when there is no energy flow across the surface.

Electromagnetic wave scattering from randomly rough surfaces

Electromagnetic wave scattering from randomly rough surfaces is fundamental to geophysics, terrestrial, and extraterrestrial remote sensing, and wireless and satellite communications. This field is a rich and active research area, and it has become impossible to cover all recent advances in analytical approaches here. The readers are referred to earlier general survey papers or books on analytical models (Elfouhaily and Guerin, 2004; Fung, 1994; Tsang et al., 1985; Ulaby et al., 1982).

The complex character of electromagnetic wave interaction with rough surface may require different classes of models in different situations. A qualitative sketch of regions of validity for analytical models is shown in Figure 1. For the region where the RMS height is very small compared to the wavelength, the small perturbation method (SPM) is applicable, which is derived rigorously from the extended boundary condition without resorting to any a priori assumption about the field. For the region where surfaces have large surface curvatures, the Kirchhoff approximation (KA) is commonly used. It is a local approximation in that it uses tangent plane approximation by assuming the surface field to be generated as if a tangent plane were at the same point while ignoring the contribution from the surface elsewhere. In the high-frequency regime, by invoking stationary phase method, the scattered amplitude reduces to a probability density function of slopes that are evaluated at specular points. In this case KA is also called the geometrical optics (GO) method. For the regions in between, there are active ongoing researches in seek of so-called unifying methods. In the survey by Elfouhaily and Guerin (Elfouhaily and Guerin, 2004), these unifying methods are classified into generic families, including the Meecham-Lysanov method, the phase perturbation method (PPM), the small slope approximation (SSA), the operator expansion method (OEM), the tilt invariant approximation (TIA), the local weight approximation (LWA), the Wiener-Hermite approach, the unified perturbation expansion (UPE), the full wave approach (FWA), the improved Green's function methods, the volumetric methods, and the integral equation method (IEM). Considering the wide use of the IEM model in the analysis of scattering from terrestrial rough surfaces on the one hand and the presentation of IEM and its related models in (Elfouhaily and Guerin, 2004) being very brief or barely touched on the other hand, these models will be elaborated a bit here to shed more light into their respective strength and limitations so as to help the practitioner make more intelligent choice.

The original IEM model (Fung, 1994) has shown to provide good predictions for forward and backward scattering coefficients. To make the derivation of the IEM model mathematically tractable, several assumptions were made, which include the following (Chen et al., 2003):



Electromagnetic Theory and Wave Propagation,

Figure 1 Qualitative sketch of regions of validity for the rough surface scattering models.

1. Spatial dependence of the local incident angle of the Fresnel reflection coefficient is removed, by either replacing it with the incident angle or the specular angle.
2. For the cross-polarization, the reflection coefficient used to compute the Kirchhoff fields is approximated by $(R_{//} - R_{\perp})/2$, where $R_{//}$ and $R_{\perp}$ are the Fresnel reflection coefficients of transverse magnetic and transverse electric waves, respectively.
3. Edge diffraction terms are excluded.
4. Complementary field coefficients are approximated by simplifying the surface Green's function and its gradient in the phase terms.

Concerns over the assumptions have prompted several modifications and variations of IEM in the literature. Regarding the spectral representation of the Green's function, the simplification was discarded and full form was restored, resulting in a modification to the complementary components. The resulted model is the so-called improved IEM model (I-IEM) (Hsieh and Fung, 1999). Additional restoration of the spectral representation of the gradient of the Green's function in its full form leads to the advanced IEM model (AIEM) (Chen et al., 2003) and the IEM2M model (Alvarez-Perez, 2001).

However, there are some technical subtleties in connection with the restoration of the full Green's function that have not been adequately reflected in these models. For example, in calculating the Kirchhoff-complementary incoherent power, one of the new quantities which need to be evaluated is the following:

$$I_F = \langle \exp\{-i[k_{sz}z + k_z z' - q|z - z'|]\} \rangle$$

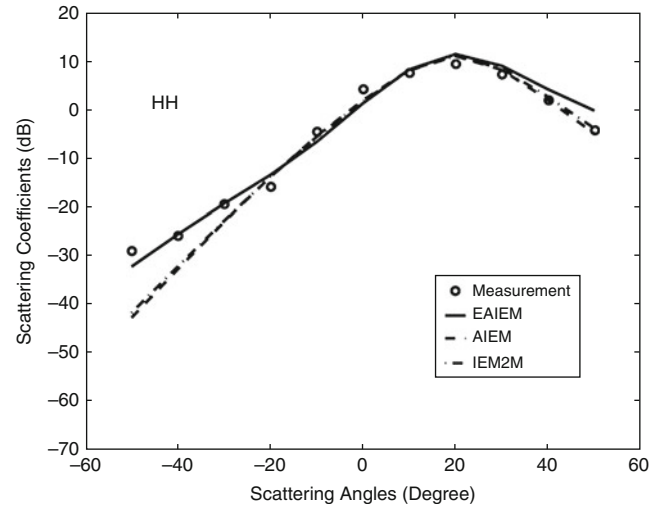
where k_{sz} and k_z are the z -components of the scattered and incident wave vectors, respectively; q is the z -related term

in the spectral representation of the Green's function; z and z' are the heights of two points on the surface, respectively, with the surface being assumed to be a Gaussian process; and the corner brackets denote the expected value. After introducing transformation of variables, one finds that I_F contains a factor of the following form:

$$I_{F2} = \frac{1}{\sqrt{2\pi\sigma\sqrt{1-\rho}}} \times \int_{-\infty}^{+\infty} dy_2 \left\{ \exp \left\{ -i \left[\frac{k_{yz}y_2}{\sqrt{2}} - \frac{k_z y_2}{\sqrt{2}} - \sqrt{2}q|y_2| \right] \right\} \times \exp \left[-\frac{y_2^2}{2\sigma^2(1-\rho)} \right] \right\},$$

where σ is the RMS height and ρ is the surface correlation function. Due to the presence of the absolute phase term, the result on I_{F2} will contain the error function. However, all the terms related to the error function have been ignored in the above models. After inclusion of the missing terms (Du, 2008), for the cross and the complementary incoherent powers, it is shown that each consists of two terms: one free from and the other related to the error function. Moreover, for the cross incoherent power, the term free from error function appears to be identical to that reported in the literature. So the other term related to error function serves as a correction term for previous calculations. It is not the case for the complementary incoherent power: The term free from error function takes a different form from that in the literature. The reason is that the evaluation of this term is more technically involved, and certain assumptions are required to make the derivation mathematically tractable. The assumptions tacitly made in Chen et al. (2003) become more explicit in Alvarez-Perez (2001). The work in Du (2008) departs from above models in another essential way by making fewer assumptions. As a result, the new model, called the enhanced advanced IEM (EAIEM), is expected to apply to a wider range with a better accuracy. This point is at least partially verified by the comparative performance (see Figure 2).

The weakness of the conventional IEM model can be examined from another angle. Assumption (1) of IEM is an oversimplification of the statistic behavior of the unit normal $\hat{n}$ directed out of the surface at an arbitrary point $\mathbf{r}$ on the surface. Specifically, $\hat{n}$ is effectively restricted to a deterministic quantity, with the direction depending on the angles of incidence and scattering, while giving up all its statistical features. Such simplification may introduce some unwanted effect on the predictive power of the model, since the statistical features of a unit normal can be rich and strong correlations can be observed between unit normal vectors. For example, consider a randomly rough surface described by a Gaussian process with Gaussian power spectrum. If $\hat{n}$ is expressed in terms of the spherical coordinates (θ_n, ϕ_n) as $\hat{n} = \hat{x} \sin \theta_n \cos \phi_n + \hat{y} \sin \theta_n \sin \phi_n + \hat{z} \cos \theta_n$, then the azimuthal angle ϕ_n and the elevation angle θ_n are



Electromagnetic Theory and Wave Propagation,

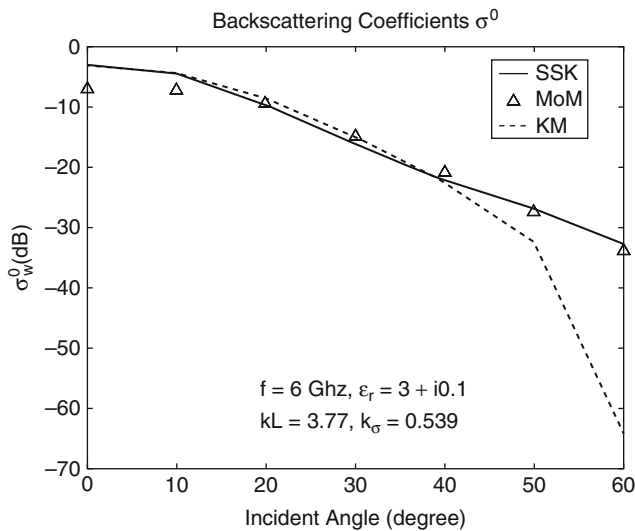
Figure 2 Comparisons of bistatic scattering models EAIEM, AIEM, and IEM2M against measurement data for a Gaussian surface. The relative permittivity of the surface is $\epsilon_r = 5.5 + i2.2$, and the frequency is 10 GHz, leading to a normalized correlation length $kL = 12.56$ and a normalized surface RMS height $k\sigma = 0.84$. The incidence angle is fixed at $\theta_i = -20^\circ$, while the scattering angle varies from -50° to 50° .

independent random variables, where ϕ_n is uniformly distributed in the interval $[0, 2\pi]$ and the probability density function of θ_n is

$$p(\theta_n) = \frac{1}{\sigma_s^2} \exp \left(-\frac{\tan^2 \theta_n}{2\sigma_s^2} \right) \frac{\tan \theta_n}{\cos^2 \theta_n},$$

where σ_s is the RMS slope. The distribution of θ_n , unlike treated in IEM, is independent of the combination of the angles of incidence and scattering. Moreover, for two-unit normal vectors $\hat{n}$ and $\hat{n}'$ at two points on the surface, respectively, if we express them in terms of the directional partial derivatives such that $\hat{n} = \frac{-\hat{x}Z_x - \hat{y}Z_y + \hat{z}}{\sqrt{1+Z_x^2+Z_y^2}}$ and form the

four dimensional vector $\mu = [Z_x, Z_y, Z'_x, Z'_y]^t$, then μ is a multivariate Gaussian random vector with zero mean and covariance matrix C . Inclusion of the statistical behaviors of the surface unit normal vectors in conjunction with the Kirchhoff approximation in the analysis of electromagnetic scattering from a randomly rough surface was presented in Du et al. (2005). The new model was called the slope statistical Kirchhoff (SSK) model. It was shown by numerical simulations that the predictions of SSK are in better agreement with MoM results than those of the conventional Kirchhoff model (KM). A typical comparison is shown in Figure 3. One finds from this figure that besides an overall better performance, SSK appears almost immune to the Brewster angle effect for vertical polarization. This feature is expected since the directions of the unit normal which lead the local angles of incidence to approach the Brewster angle occupy only a small



Electromagnetic Theory and Wave Propagation,
Figure 3 Comparisons among MoM, KM, and SSK simulations at 6 GHz with $\epsilon_r = 3 + i0.1$, correlation length $L = 3$ cm, RMS height $\sigma = 0.429$ cm.

portion of the directional distribution; contributions from the rest of the distribution become appreciable in this new treatment.

Inclusion of the statistical behaviors of the surface unit normal vectors in combination with the IEM formulism was presented in Du et al. (2007), where the shadowing effect is also included but is not expected to have any appreciable impact on the scattering behaviors predicted by the model for the range of angles of incidence considered, which is limited to be not large (say, under 70°), a restriction inherent to the IEM formulism. The new model was called the statistical IEM (SIEM) model. It was shown by numerical simulations that the predictions of SIEM are in better agreement with MoM results than those of the IEM model.

Electromagnetic wave scattering from vegetation canopy

The potential use of microwave observations to monitor vegetation water content (VWC) and soil moisture is of great importance (e.g., Storvold et al., 2006; Singh and Kathpalia, 2007). Detection of VWC is useful to monitor vegetation stress and important for irrigation management and yield forecasting. Soil moisture is often the limiting factor in transpiration of plants and evaporation from soil surface, which in turn has a significant impact on the energy cycle. Soil moisture is also a key determinant of the global carbon cycle.

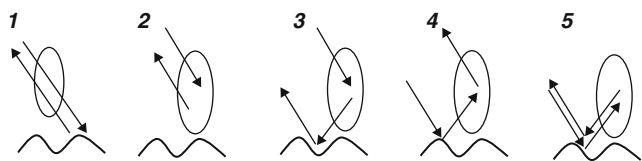
Yet to extract VWC and soil moisture from microwave observations presents a big challenge, which calls for good management of many important issues, among which are the development of a high fidelity scattering model.

In developing such scattering model, a number of factors associated with the vegetation canopy and with the underlying bare soil should be taken into account. Regarding the vegetation canopy, it is important to include the coherent effect caused by the vegetation structure at low frequencies (Zhang et al., 1995), an effect that has been well addressed in a number of recent models. For instance, the branching model due to Yueh et al. (1992) addressed the coherence effects caused by the vegetation structure, where a two-scale branching vegetation structure was used for soybeans, and the scattered fields from constituents were added coherently. A similar treatment in considering coherent effects was proposed in Lin and Sarabandi (1997) for forest canopies. Chiu and Sarabandi also considered the second-order, near-field interaction between constituents in addition to the coherence effect in their scattering model for soybean (Chiu and Sarabandi, 2000).

Yet the roughness effect of the underlying bare soil has not been adequately addressed in the above coherent models. Notarnicola and Posa, in their study of inferring VWC of corn and soybean from C- and L-band SAR images, observed that in the inversion procedure, the introduction of the dependence on roughness improves the estimates (Notarnicola and Posa, 2007). They inferred from such observation that, even for dense vegetation, the contribution from bare soil greatly influences the radar signal. To predict backscattering from the rough surface, the Kirchhoff approximation (KA) was used in Yueh et al. (1992), while a second-order small perturbation model (SPM) and a physical optic (PO) model were incorporated in Chiu and Sarabandi (2000). It is well known that the SPM model and the KA model are applicable for slightly rough surfaces and surfaces with small surface curvatures, respectively (Ulaby et al., 1982; Kong, 2005).

Regarding modeling of the vegetation, the vegetation constituents are represented by simple geometries as in Yueh et al. (1992) and Chiu and Sarabandi (2000). Specifically, stems, branches, and pods are modeled as dielectric circular cylinders of finite length. In Chiu and Sarabandi (2000) leaves are represented by elliptical thin dielectric disks, yet it was found through numerical simulation that unless ellipticity ratio is much larger than unity, the final backscatter is insensitive to the ellipticity ratio, so for simplicity leaves can be modeled as circular thin dielectric disks as in Yueh et al. (1992). The orientation distribution of the constituents is described by two angles: the elevation angle β and the azimuth angle γ , for the latter an azimuthal symmetry is assumed.

There are five major scattering mechanisms for a vegetation canopy: (1) direct backscatter from the underlying rough surface, (2) direct backscatter from soybean elements, (3) single ground bounce from scatterer to ground, (4) single ground bounce from ground to scatterer, and (5) double ground bounce. These scattering mechanisms are illustrated in Figure 4. Chiu and Sarabandi considered two additional scattering mechanisms (Chiu and Sarabandi, 2000), namely, second-order



Electromagnetic Theory and Wave Propagation,
Figure 4 Major scattering mechanisms for a vegetation canopy.

scattering interaction among vegetation constituents and scattering interaction between main stem and the rough surface. It was concluded that the latter mechanism is only considered for predicting the cross-polarized scattering at L-band (Chiu and Sarabandi, 1999). For the second-order near-field effect, good comparative illustrations with other scattering mechanisms were provided in Chiu and Sarabandi (2000) at both L- and C-band for two data sets: one from polarimetric measurements conducted using the University of Michigan polarimetric scatterometer systems (POLARSCAT) on a soybean field near Ann Arbor, MI, in August 1995, when the soybean plants were fully grown with significant numbers of pods, and the other from backscatter data collected by AIRSAR during its flight over the Kellogg Biological Station near Kalamazoo, MI, on July 12, 1995, when the soybeans were about a month old. It is observed from the decomposed contributions due to different scattering mechanisms that for not fully grown soybeans (the AIRSAR data set), contribution from rough surface is at least 25 dB higher than that from the second-order interaction for copolarized backscatter at L-band, and similar observation applies to the fully grown soybeans, so the authors concluded that the contribution from the second-order near field is negligible at L-band (Chiu and Sarabandi, 2000). At C-band, although the second-order near-field scattering is significant for fully grown soybeans as suggested by the authors, for not fully grown soybeans, contribution from rough surface is about 10 dB higher for the horizontally polarized backscatter up to 50° , while showing a cross-over with the second-order interaction for the vertically polarized backscatter at around 35° . Such observation indicates that it is more important to improve the predictive accuracy of scattering from rough surface than to include the second-order near-field effect, even at C-band, for copolarized backscatter at small to moderate incident angles. In fact, such statement may be further strengthened if one considers the fact that the estimated ground truth in Chiu and Sarabandi (2000) corresponds to a rather smooth surface, with a RMS height of 0.38 cm and a correlation length of 3.8 cm, which is several times smaller than the ground truth used in Yueh et al. (1992), where the RMS height is around 1.5 cm and the correlation length is about 13 cm. The ground truth of Yueh et al. (1992) seems to agree well with another experiment study on characterization of agricultural soil roughness for radar remote sensing (Davidson et al., 2000; De Roo et al., 2001). This suggests that the

backscatter from the rough surface might be much stronger than that of Chiu and Sarabandi (2000) if ground truth comparable to that of Yueh et al. (1992), Davidson et al. (2000), and De Roo et al. (2001) were to be used.

In Du et al. (2008) a scattering model for a soybean canopy was proposed which includes the coherent effect due to the soybean structure and takes advantage of the advanced scattering models for rough surface. Some other issues were also taken care of, such as including curvature effect in studying the ground bounce scattering mechanisms and using array theory with perturbation for characterizing the interplant structure to account for the prevailing agriculture practice of soybean.

Electromagnetic wave scattering from cylinders

Determining the electromagnetic properties of those key constituents such as branches and trunks requires knowledge of the scattering properties of dielectric cylinders (Wang et al., 2005; Henin et al., 2007; Ahmed and Naqvi, 2008). In addition, in studying scattering and absorption of electromagnetic waves from ice needles in clouds, the dielectric cylinder are also used to model those needles (Yeh et al., 1982). Thus, finding an effective method to calculate the electromagnetic scattering by dielectric finite cylinders motivated many authors. Because an exact analytical solution for the scattering from finite cylinders does not exist, several approximations have been proposed (Schiffer and Thielheim, 1979; Karam and Fung, 1988; Stiles and Sarabandi, 1996). Among them is the generalized Rayleigh-Gans (GRG) approximation, which was widely applied in the studies of the vegetation samples (Karam and Fung, 1988). It approximates the induced current in a finite cylinder by assuming infinite length. Therefore, this method is valid for a needle-shaped scatterer with radius much smaller than the wavelength. Thereafter, Stiles and Sarabandi (1996) proposed a more general solution for long and thin dielectric cylinders of arbitrary cross section, but still limited to small cross sections. Nevertheless, it should be noted that the solutions of such approximate methods in general fail to satisfy the reciprocity theorem.

In a more general setting, a semi-analytical method named T-matrix approach, originally introduced by Waterman (Waterman, 1956) and is based on the extended boundary condition method (EBCM), is one of the most powerful and widely used tools for rigorously computing volume electromagnetic scattering based on Maxwell's equations and has been applied to particles of various shapes, such as spheroids, finite cylinders, Chebyshev particles, cubes, and clusters of spheres (Mishchenko and Travis, 1994; Roussel et al., 1996; Wielaard et al., 1997). In applying extended boundary condition to calculate the T-matrix that relates the exciting field and scattered field, the exciting field is assumed to be inside the inscribing sphere and the scattered field outside the circumscribing sphere, respectively. However, for particles with extreme geometries

represented by very large aspect ratios, regular EBCM is reported to suffer from convergence problems (Barber, 1954). Physically, this ill-conditioning procedure stems from the fact that, since the exciting field is assumed to be inside the inscribing sphere, for cases of extreme geometries, the exciting fields will not be accurate representative of surface currents. Nor will the scattered fields.

One approach for overcoming the problem of numerical instability in computing the T-matrix for spheroids with large aspect ratio is the so-called iterative extended boundary condition method (IEBCM) (Iskander et al., 1983). The main feature of this technique is to represent the internal field by several subregion spherical function expansions centered along the major axis of the prolate spheroid. The contiguous subregional expansions are linked to each other by being matched in the overlapping zones. The set of unknown expanded coefficients of internal field can be determined by using the point-matching method (PMM). It has been reported that in some spheroidal cases, the use of IEBCM instead of the regular EBCM allows to more than quadruple the maximum convergent size parameter. However, because the first step in this procedure is to approximate the highly lossy dielectric object with a perfectly conducting object of the same shape for its initial solution, it is restricted by the conductivities of the dielectric particles and the maximum convergent size parameter of EBCM for such perfectly conducting object. Moreover, as pointed in (Kahnert, 2003), PMM is less flexible in terms of applications to different particle shapes due to the fact that, the more the particle's geometry departs from that of a sphere, the more unsuitable the expansions of the fields in spherical vector wave functions. Thus, elongated particles require the use of specially adapted PMM implementations with longer computation time and higher computer-code complexity. Another similar technique using PMM to solve scattering from particles enclosed by smooth surfaces is the general multipole technique (GMT), which represents electromagnetic field vectors by multiple spherical expansions about several expansion origins which are located at appropriate positions in the interior region (Al-Rizzo and Tranquilla, 1995). The GMT has been successfully used for particles with smooth surfaces, such as hemispherically or spherically capped cylinders, yet there are issues when it is used in the scattering computations of finite cylinders with flat ends reported. Recently, null field method with discrete sources (NF-DS) is proposed to deal with the instability of conventional EBCM (Doicu and Wriedt, 1999; Eremina et al., 2004; Wriedt et al., 2008). Its numerical stability is achieved at the expense of considerable increase in computer complexity, and the resolution of this method can be affected by the localization of the sources.

A new iterative technique was proposed for electromagnetic scattering by finite dielectric cylinders with large aspect ratio (Yan et al., 2008). With the understanding that for such cylinders a direct application of

the EBCM often leads to numerical instability, the procedure starts by dividing the cylinder into several identical subcylinder to reduce the aspect ratio for each part to which the EBCM can be applied. One of the technical issues that the iterative approach confronted is this: Since any two neighboring subcylinders are touching via the division interface, the conventional multiscatterer equation method is not directly applicable because it requires that the circumscribing spheres of the subcylinder exclude each other (Tsang et al., 1985). Rather, boundary conditions at the division interface need to be satisfied and carefully incorporated into the EBCM formalism. The subtlety lies in the fact that boundary conditions at the division interfaces are point-wise while the EBCM is in an integral form. For such concern, some intermediate variables that have specific meanings are introduced, where the boundary conditions are incorporated. Moreover, since these variables are expressed in terms of surface integrals, the drawbacks of PPM inherent in IEBCM or GMT are avoided. The intercoupling relations of multipole expansions for subcylinder are constructed with the help of translational addition theorems and can be solved iteratively. The impact of translational addition theorem on the convergence property of the resulting linear system is also carefully treated in the iterative procedure.

Summary

The Maxwell equations lay the foundation of all classical electromagnetic phenomena. They describe how time-varying electric fields give rise to magnetic fields and vice versa. At boundary of two different media the permittivity, permeability, and conductivity can change sharply. Boundary conditions are discontinuity equations to specify corresponding discontinuity in the field vectors. Scalar and vector potentials are useful auxiliary functions in analysis of an electromagnetic field, under Lorenz gauge condition the decoupled equations of the potentials identically satisfy the Maxwell equations. Green's function is an important tool in obtaining the solution for general source distributions. Poynting's theorem is a law of conservation of energy for the electromagnetic field. Plane waves are the simplest and fundamental electromagnetic waves. In remote sensing applications, electromagnetic wave scattering from randomly rough surfaces, from vegetation canopies, and from cylinders are important and ongoing research topics.

Bibliography

- Ahmed, S., and Naqvi, Q. A., 2008. Electromagnetic scattering from a perfect electromagnetic conductor cylinder buried in a dielectric half-space. *Progress In Electromagnetics Research*, **78**, 25.
- Al-Rizzo, H. M., and Tranquilla, J. M., 1995. Electromagnetic wave scattering by highly elongated and geometrically composite objects of large size parameters: the generalized multipole technique. *Applied Optics*, **34**, 3502.

- Alvarez-Perez, J., 2001. An extension of the IEM/IEMM surface scattering model. *Waves in Random Media*, **11**, 307.
- Barber, P. W., 1954. Resonance electromagnetic absorption by nonspherical dielectric objects. *IEEE Transactions on Microwave Theory and Techniques*, **25**, 373.
- Chen, K. S., Wu, T., Tsang, L., Li, Q., Shi, J. C., and Fung, A. K., 2003. Emission of rough surfaces calculated by the integral equation method with comparison to three-dimensional moment method simulations. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 90.
- Chiu, T., and Sarabandi, K., 1999. Electromagnetic scattering interaction between a dielectric cylinder and a slightly rough surface. *IEEE Transactions on Antennas and Propagation*, **47**, 902.
- Chiu, T., and Sarabandi, K., 2000. Electromagnetic scattering from short branching vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 911.
- Davidson, M. W. J., Le Toan, T., Mattia, F., Satalino, G., Manninen, T., and Borgeaud, M., 2000. On the characterization of agricultural soil roughness for radar remote sensing studies. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 630.
- De Roo, R. D., Du, Y., Ulaby, F. T., and Dobson, M. C. A., 2001. Semi-empirical backscattering model at L-band and C-band for a soybean canopy with soil moisture inversion. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 864.
- Doicu, A., and Wriedt, T., 1999. Calculation of the T-matrix in the null field method with discrete sources. *Journal of the Optical Society of America*, **16**, 2539.
- Du, Y., 2008. A new bistatic model for electromagnetic scattering from randomly rough surfaces. *Waves in Random and Complex Media*, **18**, 109.
- Du, Y., Xu, T., Luo, Y. L., and Kong, J. A., 2005. A statistical Kirchhoff model for EM scattering from Gaussian rough surface. *PIERS Proceedings*, **1**(2), 187–191.
- Du, Y., Kong, J. A., Yan, W. Z., Wang, Z. Y., and Peng, L., 2007. A statistical integral equation model for shadow-corrected EM scattering from a Gaussian rough surface. *IEEE Transactions on Antennas and Propagation*, **55**, 1843.
- Du, Y., Luo, Y. L., Yan, W. Z., and Kong, J. A., 2008. An electromagnetic scattering model for soybean canopy. *Progress In Electromagnetics Research*, **79**, 209.
- Elfouhaily, T. M., and Guerin, C.-A., 2004. A critical survey of approximate scattering wave theories from random rough surfaces. *Waves Random Media*, **14**, R1.
- Eremina, E., Eremin, Y., and Wriedt, T., 2004. Extension of the discrete sources method to light scattering by highly elongated finite cylinders. *Journal of Modern Optics*, **51**, 423.
- Fung, A. K., 1994. *Microwave Scattering and Emission Models and Their Applications*. Norwood, MA: Artech House.
- Harrington, R. F., 1961. *Time-Harmonic Electromagnetic Fields*. New York: McGraw-Hill.
- Henin, B. H., Elsherbeni, A. Z., and Al Sharkawy, M. H., 2007. Oblique incidence plane wave scattering from an array of circular dielectric cylinders. *Progress in Electromagnetics Research*, **68**, 261.
- Hsieh, C. Y., and Fung, A. K., 1999. Application of an extended IEM to multiple surface scattering and backscatter enhancement. *Journal of Electromagnetic Waves and Applications*, **13**, 121.
- Iskander, M., Lakhtakia, A., and Durney, C., 1983. A new procedure for improving the solution stability and extending the frequency range of the EBCM. *IEEE Transactions on Antennas and Propagation*, **31**, 317.
- Jackson, J. D., 1998. *Classical Electrodynamics*, 3rd edn. New York: Wiley.
- Kahnert, F. M., 2003. Numerical methods in electromagnetic scattering theory. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **79–80**, 755.
- Karam, M. A., and Fung, A. K., 1988. Electromagnetic wave scattering from some vegetation samples. *IEEE Transactions on Geoscience and Remote Sensing*, **26**, 799.
- Kong, J. A., 2005. *Electromagnetic Wave Theory*. Cambridge: EMW Publishing.
- Lin, Y. C., and Sarabandi, K., 1997. A Monte Carlo coherent scattering model for forest canopies using fractal generated trees. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 36.
- Mishchenko, M. I., and Travis, L. D., 1994. T-matrix computations of light scattering by large spheroidal particles. *Optics Communications*, **109**, 16.
- Notarnicola, C., and Posa, F., 2007. Inferring vegetation water content from C- and L-band SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 3165.
- Rodriguez, E., 1991. Beyond the Kirchhoff approximation. *Radio Science*, **26**, 121.
- Roussel, H., Chew, W. C., Jouvie, F., and Tabbara, W., 1996. Electromagnetic scattering from dielectric and magnetic gratings of fibers: a T-matrix solution. *Journal of Electromagnetic Waves and Applications*, **10**, 109.
- Schiffner, R., and Thielheim, K. O., 1979. Light scattering by dielectric needles and disks. *Journal of Applied Physics*, **50**, 2476.
- Singh, D., and Kathpalia, A., 2007. An efficient modeling with GA approach to retrieve soil texture, moisture and roughness from ERS-2 SAR data. *Progress In Electromagnetics Research, PIER*, **77**, 121.
- Stiles, J. M., and Sarabandi, K., 1996. A scattering model for thin dielectric cylinders of arbitrary crosssection and electrical length. *IEEE Transactions on Antennas and Propagation*, **44**, 260.
- Storvold, R., Malnes, E., Larsen, Y., Hogda, K. A., Hamran, S. E., Muller, K., and Langley, K. A., 2006. SAR remote sensing of snow parameters in Norwegian areas – Current status and future perspective. *Journal of Electromagnetic Waves and Applications*, **20**, 1751.
- Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1982. *Microwave Remote Sensing: Active and Passive*. Bedham, MA: Artech House.
- Wang, L. F., Kong, J. A., Ding, K. H., Le Toan, T., Ribbes, F., and Floury, N., 2005. Electromagnetic scattering model for rice canopy based on Monte Carlo simulation. *Progress In Electromagnetics Research*, **52**, 153.
- Waterman, P. C., 1956. Matrix formulation of electromagnetic scattering. *Proceedings of the IEEE*, **53**, 805.
- Wielaard, D. J., Mishchenko, M. I., Macke, A., and Carlson, B. E., 1997. Improved T-matrix computations for large, nonabsorbing and weakly absorbing nonspherical particles and comparison with geometrical optics approximation. *Applied Optics*, **36**, 4305.
- Wriedt, T., Schuh, R., and Doicu, A., 2008. Scattering by aggregated fibres using a multiple scattering T-matrix approach. *Particle and Particle Systems Characterization*, **25**, 74.
- Yan, W. Z., Du, Y., Liu, D. W., and Wu, B. I., 2008. EM scattering from a long dielectric circular cylinder. *Progress in Electromagnetics Research, PIER*, **85**, 39.
- Yeh, C., Woo, R., Ishimaru, A., and Armstrong, J., 1982. Scattering by single ice needles and plates at 30 GHz. *Radio Science*, **17**, 1503.
- Yueh, S. H., Kong, J. A., Jao, J. K., Shin, R. T., and Le Toan, T., 1992. Branching model for vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 390.
- Zhang, G., Tsang, L., and Chen, Z., 1995. Collective scattering effects of trees generated by stochastic Lindenmayer systems. *Microwave and Optical Technology Letters*, **11**, 107.

Cross-references

Media, Electromagnetic Characteristics

EMERGING APPLICATIONS

William Gail
Global Weather Corporation, Boulder, CO, USA

Synonyms

Emerging societal benefits

Definition

Application. An application of remote sensing is a practical use of information obtained with the technique of remote sensing. Applications are distinct from scientific investigations or engineering development in that their purpose is to consume and apply remote sensing information rather than create and understand it.

Emerging. An emerging application is one in which a new means for using remote sensing information is developed to address a new problem, exploit a new opportunity, or perform an existing function in a new way.

Introduction

Beyond its tremendous value for understanding the Earth, remote sensing also provides a myriad of more practical benefits to society (Group on Earth Observations, 2005). The National Research Council, in its landmark 2007 Decadal Survey of Earth Science and Applications (National Research Council, 2007), underscored the importance of this balance by noting that “attention to securing practical benefits for humankind plays an equal role with the quest to acquire new knowledge about the Earth system.” This balance is critical; remote sensing’s practical benefits (commonly referred to as *applications*) often arise as a direct result of the cutting-edge knowledge developed by pursuing remote sensing science.

Society’s growing need for remote sensing applications is clear. The transformation from an industrial economy to an information economy is driven by knowledge; remote sensing applications are a key element of this progress. Solutions for many of today’s critical societal needs, from health to natural disasters to the environment, require sophisticated knowledge about the Earth and its workings. The defining question addressed in this entry is simple: How will remote sensing be used by society – beyond the underlying science – in coming decades?

Classes of applications

To anticipate remote sensing applications that are likely to emerge, it is first helpful to understand today’s applications, how they are used, and what motivates their development. While these applications can be categorized in a number of ways, an effective approach is to separate them by public sector, consumer, and business uses. The boundaries between these categories may be fuzzy, but the distinctions between them are meaningful.

The *public* sector, with its obligation for ensuring the security and prosperity of the population it serves, applies

remote sensing for a wide variety of purposes. The most prevalent use of remote sensing has long been surveillance for intelligence and military use. In the civilian arena, categories of applications are often described as “societal benefits.” The most definitive list of such benefits comes from the Group on Earth Observations (GEO) and includes (a) disasters, (b) health, (c) energy, (d) climate, (e) agriculture, (f) ecosystems, (g) biodiversity, (h) water, and (i) weather. Various international and national governmental bodies (such as the United Nations Environment Programme (UNEP) and Europe’s Global Monitoring for Environment and Security (GMES)) exist to develop and use applications, but much of the applications work occurs at the regional and local governmental levels. Typical applications functions within governmental bodies include decision systems, assessments, planning, and monitoring. As with most governmental activities, new public sector applications tend to emerge through deliberate planning processes.

The *consumer* market is a growing user of remotely sensed information. The rapid adoption of consumer mapping in the last several years, both online and through personal navigation devices, provides the most tangible evidence. Remote sensing has become increasingly embedded in consumer applications, often in ways that are not obvious. Remote sensing information is now used in consumer video games, as a framework for sharing images and videos, to plan both local trips and vacation travel, to communicate news events, for exploring real estate, and much more. New applications are largely motivated by the growth of existing markets or the anticipation of new markets. In many cases, remote sensing information creates value not by being an end-use product or service itself but through enabling the purchase of other products and services (online store-locator mapping being one example).

The *business* sector employs remote sensing in two basic ways: as elements of the products and services they sell and to improve the efficiency of their business operations. Imagery from both satellites and aircraft is used to help manage operations – through functions such as geographic information systems (GIS), business intelligence, fleet management, and situational monitoring.

Context of historic trends

For much of its early history, remote sensing was dominated by the surveillance applications of the intelligence community, with an emphasis on high-resolution imaging. This information was often tightly controlled. Starting in the 1970s, the use of remote sensing for regional and global civilian applications expanded as datasets such as those supplied by Landsat and geostationary weather satellites became widely available. In the 1990s, the paradigm of Earth system science led to more sophisticated applications that began to integrate remote sensing observations from multiple sources. Finally, the rapid spread of personal computing and Internet access during the early

2000s made it far easier for businesses and consumers to access remote sensing information, triggering a wave of innovative applications. Overlaid on these four historic periods were the many evolving demands for applications, from the exploding needs of emerging economies to the globalization of commerce.

These historic trends provide a context for the ways in which applications of remote sensing have evolved. Over the last half century, applications have transformed from being a specialists' tool to a public utility, from the activities of individuals to the work of groups, from government-driven to business- and consumer-oriented, from slowly produced to rapidly distributed, from single-discipline to multidisciplinary, and from text-based to highly visual.

Current and future trends

The applications that will arise over the next several decades can be expected to reflect the classic balance between trends in user pull and technology push.

From the perspective of user pull, three applications areas are likely to be particularly important. Environmental security encompasses a number of disciplines for which remote sensing is quite useful. Environmental treaties have proliferated over the last several decades; remote sensing is critical to compliance monitoring and assessment. The growing importance of international climate agreements will only accelerate this need. Many nations are also beginning to recognize the strategic importance of the environment to their national security and economic prosperity, from preservation of water resources to understanding global demographic shifts. One important trend has been to aggregate environmental applications as a set of "ecosystem services" (Millennium Ecosystem Assessment, 2005) – a new approach that leverages market forces by managing Earth's resources as assets (similar to how businesses manage their physical assets). A second important applications area is *business efficiency*. As global markets mature, businesses seek to maintain financial margins by improving operational efficiency. Information technology, and geospatial information in particular, provides particularly high leverage for these businesses. Remote sensing applications can improve asset management, operational responsiveness, energy efficiency, and more. Third, *consumer products and services* using remote sensing are likely to grow dramatically and could substantially alter the nature of remote sensing applications. Typical consumer uses range from navigation to online commerce to gaming. The consumer market is by far the largest component of the global economy, so growth in consumer uses of remote sensing can inject substantial resources into applications development. Many important societal trends cut across all three of these applications areas, from the increasing focus on health issues to the importance of energy.

From the perspective of technology push, a wide variety of trends will shape the future. Ongoing improvements

to passive sensors, increased availability of active sensors (radars and LIDARs), and advances in sensor platforms will drive enhanced spatial and temporal resolution. Closely associated are the processing techniques used to transform data into knowledge. These range from algorithms of critical global importance, such as those designed to extract climate information from infrared radiance measurements, to seemingly trivial but high economic-value consumer image-processing algorithms such as red-eye removal. Essential to data processing advances are the ability to integrate (or fuse) data from multiple sources and automation of complex algorithms to facilitate low-cost operation across large datasets. Progress in many of these areas is also greatly enhanced by the development and use of standards, allowing easier interoperability. An often under-recognized area of sensor improvement is human-scale sensors, from professional devices for uses such as surveying to consumer sensors such as digital cameras.

Among the most important trends in supporting technology is the advent of publicly accessible satellite-based global positioning information from the US Global Positioning System (GPS), European Galileo, and Russian GLONASS satellite systems. Computing capability is also critical: Processing power continues to be characterized by Moore's Law, and similar progress metrics apply to information storage capacity and communications bandwidth. Of particular importance, the recent advances in computing and Internet access have seeded a demand for information that extends to the individual consumer. As with the development of television, this "democratization" of information is increasingly funded by advertising, ensuring that the ability to meet demand is not limited by access cost. Information access is also increasingly peer-to-peer rather than from centralized sources, with social networking and information-sharing software being the primary current drivers. The global trend toward universal availability of mobile communications devices, even within the developing world, greatly facilitates this progress. Finally, widespread access to digital photography (in part through inclusion of cameras on mobile telephones) is making people more familiar with the uses of imagery and creating a class of information that can augment and enrich the traditional government-owned sources of remote sensing information.

The rate at which new remote sensing applications emerge based on these trends will also be modulated by larger societal "megatrends." Crisis situations, from military conflict to long-term natural disasters, tend to be strong motivators for rapid advances in remote sensing. The global response to climate change and the international competition for energy may be such drivers; others that are more difficult to anticipate are likely to arise as well. Major changes or opportunities in global economic markets are a second important large trend. The market transformation created by the Internet and the shift of economic growth toward developing nations are important examples.

Characteristics of emerging applications

Emerging remote sensing applications are typically characterized by one or more of the following general attributes:

- *New uses for existing remote sensing capabilities.* Examples include migration of established remote sensing techniques to new markets and new public sector uses arising from newly created programs or agencies.
- *New sensors or techniques which enable applications that were previously technologically, economically, or functionally prohibitive.* In many cases, the applications needs of governments and businesses are well known, but they are difficult or too expensive to address until technology advances make them feasible.
- *New ways of analyzing, processing, or combining remote sensing information.* Useful applications often require extensive algorithmic processing of raw data. Progress is enabled by algorithm development and advances in computing power.
- *New science that enables new uses.* Scientific investigations using remote sensing routinely expand our knowledge of the Earth. This knowledge is used to enhance existing applications and to motivate development of entirely new applications.
- *Technological advances in related areas.* Many new applications emerge as a result of collateral technology advances in computing, communications, and satellite navigation.

These general attributes provide the framework for specific characteristics of emerging applications that can be anticipated over the coming decade and longer:

- *Real-time applications.* The need for remote sensing applications is often time critical, yet access to sensor information and the required processing can take hours or days. Satellite providers have recognized this and are emphasizing reduced revisit times for low-Earth-orbit constellations, radar sensors that can image at night, geostationary platforms to provide persistent imaging, and rapid deployment of aerial platforms (including drones). Advances in automated processing have also been critical to achieving rapid information dissemination.
- *Multidisciplinary integration.* Simple applications tend to focus on a specific need or single information source. The maturing of remote sensing has enabled development of applications that begin to connect multiple information sources. This can involve data from multiple scientific disciplines (e.g., weather and oceans) as well as diverse types of information ranging from science and business.
- *Shared and multiuser applications.* Until recently, most web-based software followed the server–client architecture, in which a centralized server publishes information to multiple users. Increasingly, web-based software employs a peer-to-peer architecture

(commonly referred to as Web 2.0 and social networking) that facilitates direct sharing of information between individual users. Similarly, applications have begun to support multiple simultaneous users, enabling shared viewing of dynamic content and collaborative editing of information.

- *Transition from text to visualization.* Visualization is a highly effective means for communicating large amounts of information, particularly when spatial information is involved. Extensive visualization within applications requires systems with large bandwidth and considerable data storage, capabilities that have become widely available within the last few years.
- *Integration across spatial/temporal scales.* Applications have historically been constrained in spatial and temporal scale: limited to a global, regional, or local use and focused on either long-term or short-term. Applications that support linkages between scales are rapidly emerging. A prominent example is medium-range [weather forecasting](#) that integrates multidecadal global climatology, regional/seasonal patterns such as El Niño, and localized hourly weather forecasting.
- *Improved data archive and access.* As information volumes have grown, the ability to archive, find, and rapidly access information has become a significant challenge. Considerable resources are being applied to this problem through web-based search, including the areas of metadata, data tagging, indexing, and vertical search. With increasing consumer use, the business model for accessing remote sensing information has evolved; instead of end-users paying for access to the basic data, revenue increasingly comes from use of the applications that consume this data, including “free” applications monetized through advertising revenue.
- *Rapid, widespread communications.* Effective applications quite often require the rapid and widespread dissemination of information, decisions, and alerts. The ongoing revolution in mobile communications, both networks and handsets, has opened many new opportunities for remote sensing applications. Coupled with the Internet, mobile communications enable broad global information sharing and ready access at the point of use.
- *Merging of real and hypothetical information.* The transition from analog to digital information has made possible extensive processing and reprocessing of information. Within this context, it has become easy to combine or merge real and hypothetical data, and this has proven valuable for remote sensing applications. Simple examples are the 3D visual representation of a city modified to include a hypothetical development project and the visualization of hypothetical ecosystem changes brought on by climate change.

Anticipated applications

Based on these trends and characteristics, it is possible to anticipate several important classes of applications for the next decade and beyond. *Environmental services*, both

for government and business use, will almost certainly lead the way. These will encompass enhancements to existing applications, such as weather forecasting and resource management, as well as new applications to support emerging needs in the areas of climate, energy, air quality, and health. Society's increasing sensitivity to human and property losses from natural disasters will motivate improved applications addressing *disaster prevention and response*, with emphasis on communicating centralized information and decisions to the general population. The private sector will increasingly integrate remote sensing and other geospatial information into their operations to improve *business competitiveness and efficiency*. Both businesses and consumers will benefit from the growing availability of *location-based services* and augmented reality, allowing them to get information, make decisions, and perform functions specifically tailored to where they are or where they are going. Associated with this will be significant advances in the capabilities of *personalized mapping*, the ultimate tool for getting remote sensing out to the individual, including both Internet and mobile uses. Finally, *gaming and entertainment* cannot be underestimated due to the considerable financial resources available within this market; innovative uses of remotely sensed information will be developed to enhance the realism of applications in these areas.

It is also possible to anticipate some of the more innovative applications that may emerge in coming decades. High on this list are applications that involve *community-based mapping and remote sensing*. The ability of government and businesses to monitor the Earth at all space and time scales is currently highly constrained. Community-based data gathering (with sensors such as mobile phone cameras), and community-based applications, will be increasingly important. Various technological advances, from the proliferation of mobile phones to the spread of social networking software, make this possible. Linkages between these community-based networks and more formal government/business networks will leverage the capabilities of each. As the world becomes increasingly monitored across all space and time scales, this knowledge will further our ability to create replicas of the real world – 3D *virtual worlds* in which many of the real world's functions can be mirrored. These worlds – built and updated with remote sensing information – provide a virtual context in which applications can be visualized and applied without the constraints encountered in the real world. Perhaps most exciting, they create a platform for “what if” applications, allowing us to implement functionality (“experience a tsunami”) and perform actions (“assess the societal consequences of a global crop blight”) that cannot be accomplished within the real world for safety, ethics, or other considerations.

Summary

Applications of remote sensing have proven, over the last several decades, to have tremendous benefit to

society – increasing our prosperity and making us more secure. The need for such applications is only accelerating as we transform from an industrial society to one that is information-driven. Advances across all areas of remote sensing, as well as the enabling computing and information technologies, ensure that emerging applications of remote sensing will be increasingly important to society's progress.

Bibliography

- Group on Earth Observations, 2005. *Global Earth Observations System of Systems GEOSS: 10-Year Implementation Plan Reference Document*. Noordwijk: ESA Publications Division.
- Millennium Ecosystem Assessment, 2005. *Living Beyond our Means: Natural Assets and Human Well-Being*. Washington, DC: Island Press.
- National Research Council, 2007. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. Washington, DC: The National Academies Press.

Cross-references

[Cost Benefit Assessment](#)
[Emerging Technologies](#)
[Environmental Treaties](#)
[Policies and Economics](#)

EMERGING TECHNOLOGIES

Jason Hyon

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

Active sensing; Data Management; Instrument; Passive sensing; Technology trends

Definition

Technologies to enable measuring new science requirements.

Emerging technologies

The objective of the *emerging* technology area is to develop highly capable instruments that are a key element in contributing to an ability to successfully develop and operate high-performance future remote sensing missions. The current state of the art remote sensing sciences tend to demand higher special and spectral resolution from next generation remote sensing capabilities. With a growing emphasis on cost-effective approaches due to high spacecraft and launch vehicle costs, multiband and low mass sensors are of key interests in this area.

There are a number of high payoff technologies within this area that include active remote sensing, massive data processing techniques, and technology push. The primary technology challenges for active remote sensing is

development of systems in ever-broadening envelopes of transmission; large aperture, light-weight, high-efficiency, and high-power transmitters; and reliable systems. These technologies will enable [RADAR](#) and [LIDAR](#) for a consistent coverage of the globe for CO₂, biomass, cryosphere, water cycle, and earthquake (Richard et al., 2007).

There is a robust and mature capability that is in place for Earth science enterprise to understand surface, ocean, and atmospheric characterizations of Earth. There have been many high resolution instruments to measure various parameters to study Earth such as [RADAR](#), [LIDAR](#), hyperspectral imager, and spectrometer instruments. With use of distributed data architecture, Earth science enterprise distributes data based on middleware architecture and provides a framework for data mining, modeling, simulation, and visualization capability. With significant increase in data volumes, traditional methodology for data processing will not be able to meet challenges of high-throughput modeling and data extraction. These developments should be incorporated to provide a uniform interface to the operational team. Handling large volume of data and information generation to decision support should take advantage of advancement from GIS and industry de facto standards.

In addition to the high-capability technology areas that are discussed, it is important to note the technologies that promise the future. These are the innovative technologies that have the potential to enable critical new science or exploration capabilities. These are promising because they do one or more of the following: stretch the state of the art in some respect, have the potential for high performance, provide a novel operational concept, and attempt to perform a measurement in way never before envisioned. For technology push areas, the technologies required for these instruments focus primarily on detector technology, whether in the detector sensitivity, size of detector arrays, or detection stability and calibration. For some applications, large apertures are also a challenge. In the areas of in situ sensing, challenges associated with in situ sensors focus on development of sensor webs that produce science-lab quality measurements in a field instrument. Furthermore, high-capability instruments tend to produce a lot of data volume and would require high downlink capability; optical communication will play an important role in the future.

Summary

Future science requirement demands higher resolution and temporal coverage. Active sensing techniques and large focal plane array detection will enable those objectives along with communication and information technology infrastructure.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Richard, A., Berrien, M., II, et al., 2007. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. Washington, DC: The National Academies Press.

Cross-references

[Data Access](#)
[Data Archival and Distribution](#)
[Emerging Technologies, LIDAR](#)
[Emerging Technologies, Radar](#)
[Emerging Technologies, Radiometer](#)
[Emerging Technologies, Sensor Web](#)
[Microwave Horn Antennas](#)
[Reflector Antennas](#)
[Thermal Radiation Sensors \(Emitted\)](#)

EMERGING TECHNOLOGIES, FREE-SPACE OPTICAL COMMUNICATIONS

Hamid Hemmati

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Introduction

Future pursuit of the vision for robotic and human space exploration would utilize instruments with ever-increasing capability and require orders of magnitude increase in data return rates from planetary distances. Without resorting to high-power transmitters with hundreds of watts and antenna diameters on the order of 5 m, deep-space science and exploration will rapidly encounter a bandwidth ceiling with radio frequencies of S, X, and Ka bands. In addition, communicating at a given data rate with an outer planet such as Neptune or Pluto is 1,000 times more difficult (when considering link margin degradation) than communicating at nominal Mars distance. Similarly, a link from Mars is approximately 10 billion times more challenging than communicating from geosynchronous orbit (GEO) to the ground (Hemmati, 2006). On space platforms, mass and electrical power are expensive commodities, and thermal gradients and ionizing radiation pose challenges as well (Hemmati, 2009).

Laser communications (lasercom or optical communications) through free-space, utilizing a narrow laser beam transmitted from planetary distances, hold great promise for delivering the increased data rates while imposing reasonable mass and power burdens on a host spacecraft. This technology has the potential to enable deep-space communication between Earth and planets at the farthest reaches of our solar system and beyond. In the coming decade, with comparable mass and power of a radio-frequency communication system (including optics one tenth the diameter of the flight antenna aperture), lasercom systems are expected to deliver at least an order of magnitude higher

data rate than conventional radio-frequency systems. Lasercom could offer these benefits with no additional attitude-control burden on the spacecraft and any spectrum bandwidth allocation or restrictive frequency regulations.

The high-capacity communications offered by lasercom could enable scientists to communicate information from other planets using, for example, streaming HDTV video, synthetic aperture radar, multi- and hyperspectral imagers, precision navigation, and multichannel telecommunications. Additional possibilities are multifunctionality with space-based optical navigations, laser remote sensing, imaging, and altimetry (Hemmati and Lesh, 1998).

Recent highly successful in-space validations of lasercom technology (e.g., GOLD, GeoLITE, SILEX, and TerraSAR-X) have led to the development of operational systems (Wilson et al., 1997; Caplan, 2007; Tolker et al., 2002; Smutny et al., 2009). Although lasercom has been fairly well proven in the Earth's orbit, it will need to undergo a number of successful deep-space demonstrations – with an order of magnitude higher data rate and at a lower cost per bit than conventional systems – before it can be operationally implemented in space. To date, there have been no laser telecommunications with a planetary spacecraft, although there have been demonstrations of aiming a ground-based laser at spacecraft into deep space, and vice versa (Wilson et al., 1993; Smith et al., 2006).

An introduction to significant system design and system engineering aspects of the flight and ground portion of a planetary laser communications will follow.

System design and engineering

Telecommunication at planetary distances involves link ranges that can range up to billions of kilometers. Due to the $1/r^2$ dependence for the space loss factor, a system engineer faces the challenge of designing a system that can compensate for an additional loss of 60 to greater than 100 dB over that of a near-earth satellite-to-ground link. A clear improvement area is to increase the aperture diameter at both ends of the link (Hemmati et al., 2007). However, an increased flight transceiver aperture exasperates the laser beam pointing challenge, since efficient implementation necessitates beam pointing to the order of one tenth of the telescope's beamwidth. Moreover, larger ground apertures collect larger amounts of background light, and even the highest quality telescopes are constrained by atmospheric turbulence-induced spreading of the signal (blur circle) at the telescope's focal plane. As discussed below, before a system can be designed, the link requirements and parameters have to be considered in its entirety.

Constraints and desired characteristics of an optical link

Although optical communication systems offer many potential advantages in deep data return capability, the current implementation also has a number of limitations that can affect the link availability and overall data return strategy. These constraints include limited weather-related

availability from a given ground station and limited solar conjunction availability arising from Sun platform geometry. Also, until the spacecraft/mission design can provide sufficient confidence in autonomous systems, radio-frequency (RF) link coverage (albeit with reduced capability) will likely augment the optical link during the attitude-constrained mission phases and critical maneuver.

Desired characteristics of an optical link include (Hemmati, 2006):

- Support high data rates that are competitive to RF implementation
- Support bidirectional links and spacecraft navigation
- Provide high reliability
- Low mass, power consumption, and volume
- Radiation tolerant
- Not to place demands on the host platform
- Provide extended mission coverage
- Function at small Sun angles

Major design driving issues include:

- Stabilization and pointing of the narrow laser beam transmitted from the remote platform
- Minimization of scattered light, thermally induced telescope deformations, and avoiding telescope damage – considering that transceivers at each end of a link must typically look near the Sun in order not to avoid operational downtime
- Rejection of as much of the collected background light as is feasible (e.g., via optical filtering) prior to data detection

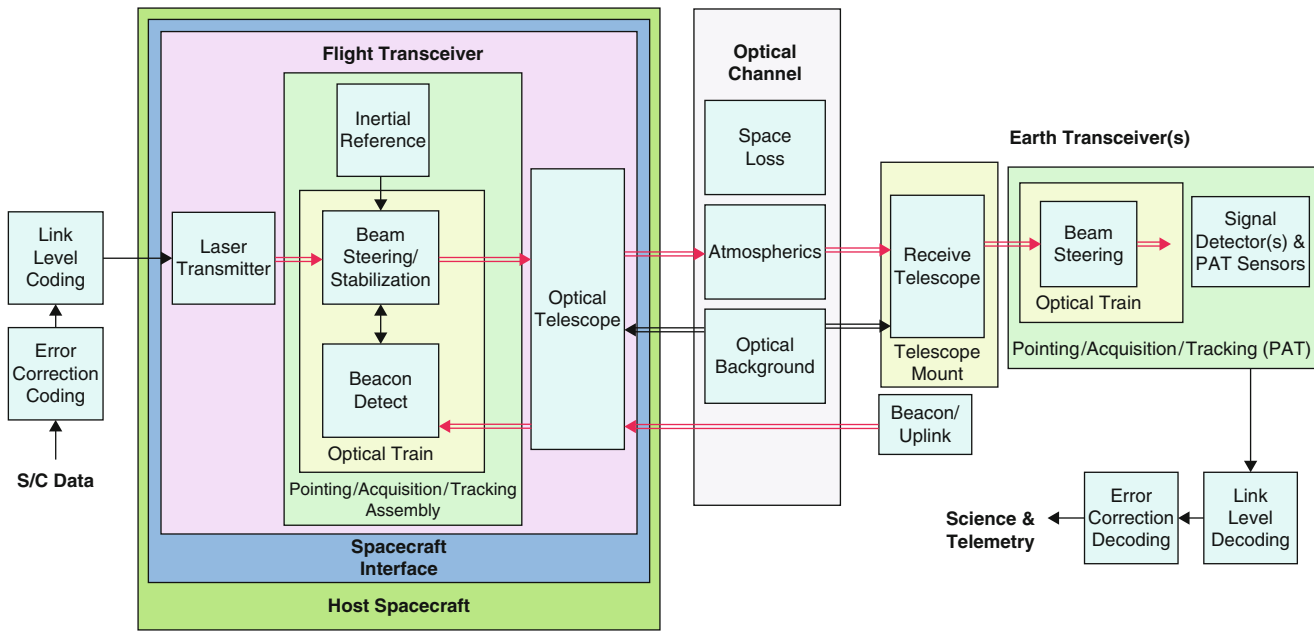
End-to-end link block diagram

A top-level, end-to-end signal flow for an optical communications signal flow is depicted in Figure 1. Mismatch among various blocks or performance degradation in any subsystem block, for example, due to device aging or radiation damage, will result in reduced efficiency of data transmission. The host platform random vibrations, the space platform/transceiver interface, and the space environment (radiation and vacuum) can all influence the flight transceiver's behavior.

Communications link design

The link design control table (DCT) is a listing of link influencing design parameters, such as aperture size, output power, and other requirements placed on the transmit and receive transceivers, and the resulting estimated system at a specific point in time during the mission (Hemmati, 2006). DCT analysis, with the primary purpose of estimating the required signal power to be delivered to the opposite terminal and available margin to maintain the desired link performance for a given worst-case condition, is a critical consideration in the design of a laser communications system.

Factors driving the link analysis are (1) fixed requirements/specifications (e.g., link distance, data rate, and bit error rate) provided by the user, (2) those fixed by laws of physics (i.e., space loss), (3) variables of the hardware



Emerging Technologies, Free-Space Optical Communications, Figure 1 Telecommunications signal flowing from a spacecraft probe to Earth.

design (e.g., aperture sizes and laser powers at opposite ends of the link), and (4) variable hardware qualities driven by implementation approach (e.g., pointing losses and optical losses). By adding the gain factors together (e.g., laser power, photodetector sensitivity, and aperture's area – in units of dBm or dB) and subtracting from the loss factors (e.g., the space loss, mispointing, and imperfection of optics), we are left with the link margin that needs to be positive in order to establish a link.

The mean received signal power at the receiver is related to the transmit power and various link losses, and the transmit and receive antenna gains can be estimated from the following equation:

$$P_R = P_T \eta_T G_T L_{TP} L_{atm} L_s G_R \eta_R$$

Where

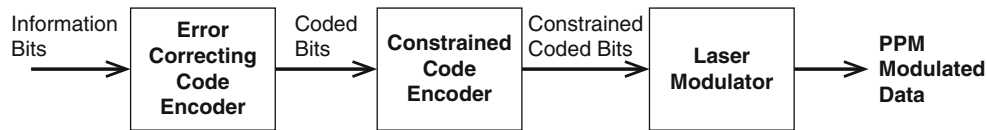
P_R	Signal power incident on the data detector
P_T	Output power of the laser transmitter
η_T	Transmit optics efficiency, accounting for relay optics transmission, wave front quality errors, and beam truncation
G_T	On-axis transmitter gain
L_{TP}	Transmitter pointing loss
L_s	Space loss
L_{atm}	Atmospheric loss comprised of attenuation and turbulence (beam spreading beam jitter losses)
G_R	Receiver gain
η_R	Receiving efficiency accounting for optical transmission and losses associated with mispointing and detector overfilling or truncation (Hemmati, 2006)

Technologies for data rate generation and detection

In this section, trades on data modulation and reception techniques, coding, laser transmitters, single-photon detectors, decoding, demodulation, and data-receive electronics are discussed.

What detection scheme we use is intertwined with what data transmission format we employ, and vice versa. Both direct detection and coherent heterodyne reception are options for data reception. A direct-detection receiver detects the intensity of the received signal and (without background interference) can achieve performance exceeding that of heterodyne detection. Implementation of narrow field-of-view optical receivers and narrow band-pass filters in the receiver chain becomes essential to mitigate the background noise and improve receiver sensitivity. Coherent (heterodyne) receivers first interfere with the downlink optical signal via a local oscillator (laser) prior to photodetecting the combined signal. The signal beam mixing process is mode-selective spatially, and only the signal energy that is mode matched to the local oscillator beam is detected. Therefore, the coherent detection scheme is largely immune to background light noise. However, the coherent receiving process is significantly more complex, particularly for a downlink signal that has traversed the turbulent atmosphere.

Considering the photon-starved regime of operations at ranges typical of planetary distances, our overall goal is for the most efficient detection (measured by the



Emerging Technologies, Free-Space Optical Communications, Figure 2 Example of a coded transmit optical channel, modulated in PPM format.

unit of bits/photon). Coherent receivers utilizing either homodyne techniques or heterodyne techniques with coherent IF electronics have achieved as much as -3 dB bits/photon (for homodyne-detected) signals (Borson, 2007). The receiver sensitivity for (typically) photon-starved regime planetary communications links is of critical importance. The required receiver sensitivity is on the order of a few photons/bit for expected telemetry downlink communications rates, ranging from 10 to 100 s of megabits/s.

Phase fluctuations (arrival angle variations) constitute the primary influence of turbulent atmosphere on a direct-detection receiver since, by utilizing large (multimeter) diameter ground-based telescopes, the scintillation effect is averaged out to a great extent. In direct detection, the atmospheric turbulence limits the phase coherence across the telescope's optical aperture to a distance of r_0 , known as the Fried's atmospheric coherent length (Andrews and Phillips, 2005). The value of r_0 is highly variable, depending on the receiver site and the time of day, and can vary from a few centimeters (daytime – poor seeing conditions) to tens of centimeters (nighttime – excellent seeing conditions). For distances up to r_0 , arrival angle/phase remains mostly correlated. Once the daytime adaptive optics technology is sufficiently advanced, a substantial portion of the turbulence effect on the lasercom signal can be mitigated.

Currently, coherent detection approaches require single-spatial-mode reception. However, propagation of the laser beam through the turbulent atmosphere results in distorted (non-flat) wave fronts, leading to an enlarged blurred circle (multimode beam) at the telescope's focal plane. In the absence of an often elaborate adaptive optics system for a large ground-based telescope, the multimode beam is unsuitable for both coherent mixing with local oscillators and for coupling to single-mode fibers. Implementation of adaptive optics during daytime and meeting the unique requirements imposed by lasercom on adaptive optics (for generation of very high Strehl ratios – approaching 1.0) require additional technology development (Wilson et al., 2003). In the absence of an atmosphere (e.g., cross-links between assets above the atmosphere), coherent detection would provide a highly competitive performance.

For a turbulence-constrained system, the simplest detector is one that provides what is known as “direct detection.” The most efficient detector of this kind is sensitive to single photons and counts individual photons with high-detection efficiency and extremely low noise (discussed in further detail later) (Farr, 2009). For this detection scheme, the

requirement for spatial coherence is relaxed to levels sufficient for focusing a received beam onto the active area of the photodetector, located at the focal point of the optical aperture, and collecting the received photons.

Modulation and coding

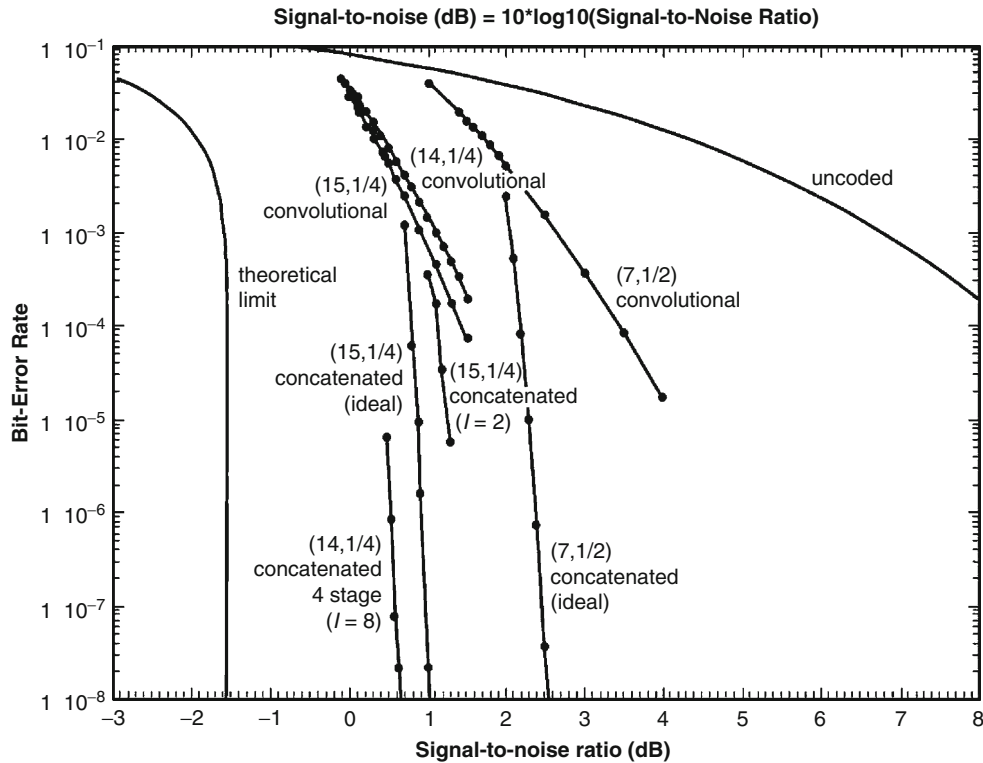
Figure 2 shows an example of signal flow through a coded free-space optical channel. Coding gains exceeding 5 dB can now be expected from the advanced high-efficiency codes, approaching to a fraction of a dB of the Shannon capacity limit (Hamkins and Moision, 2004). The resulting link efficiency gains have the direct effect of relaxing requirements on the transmitted laser power and telescope's aperture diameter, resulting in an overall simplification of the flight transceiver.

One may model a direct-detection free-space optical link as a Poisson point process, with system limitations on bandwidth, average power, and peak power (Moision et al., 2006). Under these constraints, a negligible capacity loss is encountered when confining the modulation to a binary, slotted format (Moision, 1999). The capacity of an optical channel, which represents the highest data rate that the channel can consistently support, depends on multiple factors including the modulation format, the medium through which the beam traverses, the optical preamplifier (if any), and the photodetector. A channel error-correction code (ECC) may be found with negligible output probability of error, when data rate is less than a given channel's capacity. However, the probability of error is enhanced when the data rate exceeds the channel capacity (Wyner, 1988).

Most efficient optical links result when the received signal's peak-to-average power ratio is high (Hamkins and Moision, 2004; Lipes, 1980). M-ary pulse-position modulation (PPM) is a modulation scheme that can create such a condition – resulting in near-capacity performance and negligible loss relative to alternatives. In the PPM format, $\log_2 M$ bits choose the location of a single-pulsed slot in an M-slot frame (Hamkins, 2008). Error-correction codes based on *serially concatenated codes* have been designed for the PPM channel, where the constituent codes are:

- Convolutional code serially concatenated with PPM (and preceded by a bit accumulator)
- Low-density parity-check (LDPC) code concatenated with PPM (Tan et al., 2008)

These codes – when used in conjunction with photon-counting detectors – achieve efficiencies better than 1 dB photons/bit. Implementation of PPM does result in



Emerging Technologies, Free-Space Optical Communications, Figure 3 Efficiency of state-of-the-art optical codes compared with channel capacity.

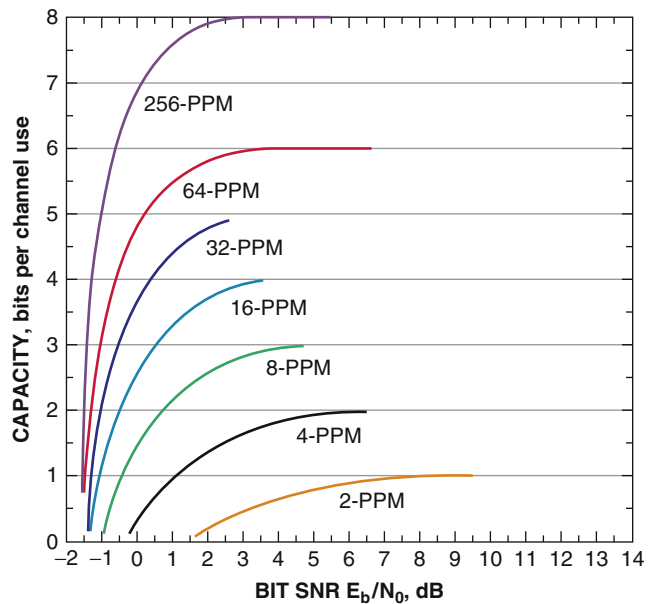
finite losses (due to modulator constraints) and more importantly necessitates high (kW level)-peak-power lasers.

Figure 3 shows the power efficiencies available at various data rates, using current technology of modulation, coding, and reception. On-off keying (OOK) modulation in conjunction with low-noise optically preamplified receivers is a common approach for fiber-optic telecommunications at rates exceeding 40 Gb/s. Utilizing the best available decoder chips, in conjunction with Shannon capacity-achieving codes (e.g., turbo codes) and soft decisions, efficiencies on the order of 6 dB photons per bit (at 10^{-9} BER) have been achieved (Kiasaleh, 1998).

Figure 4 shows the efficiency of PPM at differing alphabet size, M . This figure illustrates how one may trade bandwidth for efficiency by using higher and higher M , and how a noisy environment significantly degrades the performance efficiency of a photon-counted PPM-modulated signal (every 3 dB of extra noise requires about 1 dB of extra signal) (Dolinar et al., 2000). The difference between a noiseless and just one noise count/slot can be as much as 10 dB of efficiency, diminishing the significant advantages provided by the photon-counting technique (Moision and Hamkins, 2003).

Laser transmitter

Implementation of high-order PPM requires laser transmitters with a high peak-to-average power ratio.



Emerging Technologies, Free-Space Optical Communications, Figure 4 Efficiency of different PPM orders indicating that the recently developed codes for optical detection are within a fraction of 1 dB of capacity limit. PPM capacity on the additive white Gaussian noise (AWGN) channels is compared with other performance curves.

Emerging Technologies, Free-Space Optical Communications, Table 1 Current status of key single-photon-sensitive detectors

Detector	Wavelength (nm)	Detection efficiency	Dark rate (Hz/mm ²)	Notes
Geiger-mode APD	400–850 900–1,600	40 % 55 %	1E6 2E4	High jitter. Reset time with each detection event
Intensified hybrid photodiodes	900–1,600	40 %	1E6	Photocathode degradation
Superconducting	400 to > 1,600	80 %	1E2	3°K temperature operation
Negative avalanche feedback	450–800 900–1,600	35 % 11 %	3E4 5E8	Intolerant of over bias. Low maturity

Master-oscillator power-amplifier (MOPA) sources based on Yb-doped fiber (operating at 1,060 nm) or Er/Yb-doped fiber (operating near 1,550 nm) satisfy this requirement and are compact enough for spacecraft use (Hemmati, 2009). Modulation extinction ratios on the order of 80 dB have been demonstrated with picosecond lasers (Braun et al., 1995). Also, mode-locked fiber lasers with 1 GHz pulse repetition rate and over 60 dB secondary mode suppression have been achieved (Deng et al., 2004).

Utilization of the laser in deep space also necessitates an overall high (input electrical to optical) efficiency of operation. A wall plug oscillator/amplifier efficiency of greater than 30 % is theoretically achievable (Hemmati et al., 2000). Currently, 1,060 nm lasers are proving to be more efficient than their 1,550 nm counter parts (Wysocki et al., 2006; Spellmeyer et al., 2005). MOPAs modulated at hundreds of megahertz with average output powers exceeding 10 W, peak powers approaching 1 kW, and sub-ns pulses are now available (Dupriez et al., 2006). The performance of current fiber amplifiers is limited by self-phase modulation (SPM) and to a lesser extent by stimulated Brillouin scattering (SBS) nonlinear effects when pulse width is <0.1 ns (Charplyvy and Tkach, 1993).

Laser designers also emphasize suppression of undesired background radiation noise, known as amplified spontaneous emission (ASE), which is generated along with the amplified signal (Bromage et al., 2003). ASE's spectrum spans that of the amplifier's gain profile and maintains the same polarization as the amplified signal. This can cause transmit/receive isolation challenges for the transceiver. Winzer et al. show that for near-Earth ranges of communications, the ASE noise levels may rival the background radiation received from the Sun (Winzer et al., 1999).

Uplink laser(s) for transmitting data to spacecraft and/or as a beacon (to assist with the acquisition and tracking of the earth station by the remote transceiver) typically requires orders of magnitude higher power and orders of magnitude lower pulse repetition frequency (Biswas et al., 2005). Unlike the downlink laser, sub-nanosecond pulse widths will not be required for the uplink laser.

Single-photon-sensitive photodetectors

Efficient detectors at each end of the optical communications link reduce the required laser power at the opposite end of the link. Top-level photodetector requirements,

when employing PPM modulation in conjunction with photon-counting receivers, include high single-photon detection efficiency, low output pulse timing jitter, low reset time, large active area simultaneously with high bandwidth, and low dark count rate (Robinson et al., 2005).

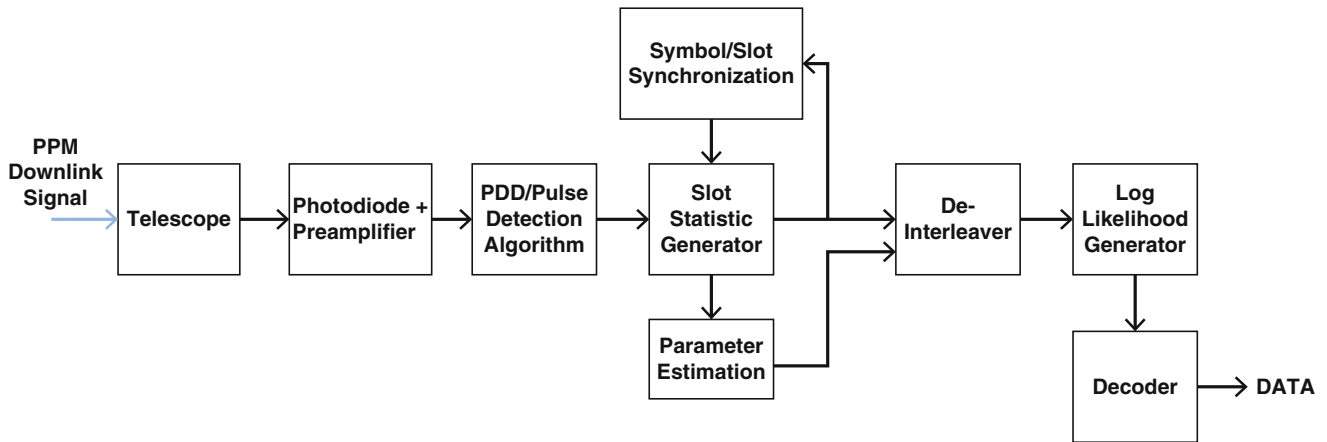
The advent of new generations of ultra-low-noise, high-gain, and high-bandwidth single-photon-sensitive detectors (also known as photon counters) has resulted in nearly 8 dB of link margin gain relative to APDs, enabling optical communication links with sensitivities approaching one photon/bit (Farr, 2009; Stern and Farr, 2007). When photon counters are utilized along with near optimal (channel capacity approaching) codes and high peak-to-average power lasers, nearly 20 dB gain may be realized relative to direct-detection systems operating with on-off keying (OOK) modulation (Borson, 2007). Minute levels of detector noise also enable implementation of an array of optical receivers each equipped with a photon counter, followed by electrical signal summation (Quirk and Gin, 2006). Table 1 compares current merits of leading photon counters.

The *reset time*, an imperfection that is characteristic of most photon counters, is the period of time for which the detector is inactive upon each signal or noise detection event (Kerman et al., 2006). Depending on the detector type, the reset time period may span from a few nanoseconds to a few milliseconds. This drawback limits the achievable data rate and the total detectable photon flux (detection event/s).

For detection of very faint signals and to avoid detector shutdown due to dark count events, it is advantageous if the mean time between dark count (noise) signals is shorter than the reset time (Kachelmyer and Borson, 2007). In a photon-starved regime (such as planetary links), daytime background light flux exceeds dark count rate of the detector. The minimum resolvable arrival time for a photodetection event is known as the output pulse *timing jitter* of a given detector (Moision, 2008a). This timing resolution restricts the number of bits accessible from each detection event, setting an upper limit on the achievable data rate (Moision and Farr, 2008).

Receive electronics

Various architectures have been envisioned for data-receive (synchronization, demodulator, decoder) electronics. In one *distributed and scalable* photon-counting



Emerging Technologies, Free-Space Optical Communications, Figure 5 Schematic block diagram of a generic PPM receiver.

decoder architecture, a receiver that independently synchronizes to the received signal follows each detector (Gin et al., 2007). Sets of pulse-position modulation (PPM) slot statistics generated by multiple receivers are combined in each slot accumulator and fed to a decoder for recovering clock and transmitted data bits. Multiple decoders will be required for high (tens of megabytes per second) data rate links. In distributed decoder architecture, each unit that follows the receiver contains a single PPM decoder. A decoder that is unavailable relays the undecided code words onto subsequent decoders. The receivers, slot accumulators, and decoder boards are connected using multi-gigabit serial deserializers. The distributed decoder and slot accumulator boards may share a common hardware platform but differently programmed field-programmable gate arrays. The scalability feature of this architecture has been utilized to develop receiver electronics for planetary-type links utilizing high-order PPM demodulation/decoding and capable of handling hundreds of megabytes per second of data rate throughput (Moision, 2008b). Figure 5 depicts a generic block diagram of a PPM receiver, including synchronization, demodulation, and decoding of the photodetected signal.

Optics for flight transceiver

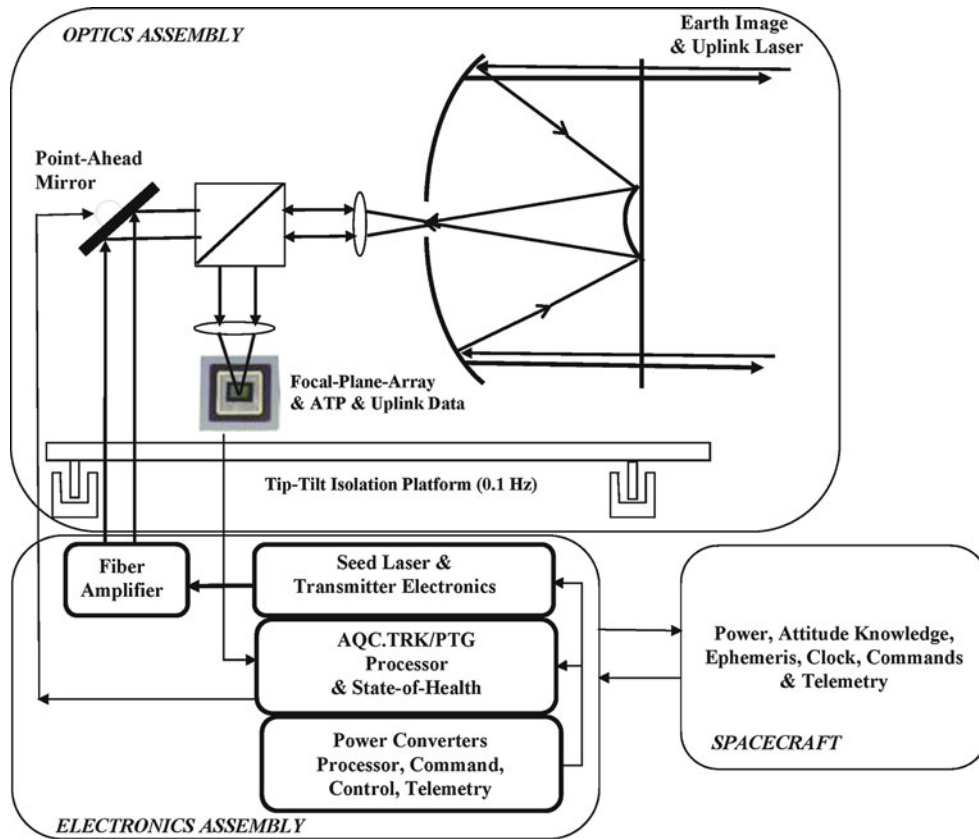
Figure 6 shows a generic schematic diagram of the flight optomechanical assembly. This subsystem is comprised of three primary channels: laser beam transmit, data receive, and pointing, acquisition, and tracking. The laser transmitter assembly (typically located remotely to simplify thermal management) is connected to the optics assembly via a single-mode fiber. The data-receive channel may include an avalanche photodiode detector (APD) or a single-photon-sensitive detector. Synchronizer, demodulator, decoder, and amplifiers assisting with

clock and data recovery comprise the remaining receiver electronics.

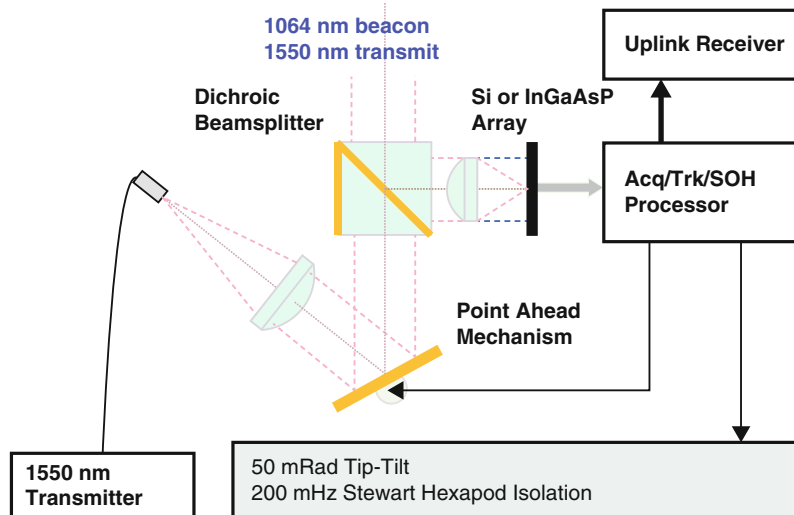
Channeling the uplink signal to the data receiver may be accomplished through free-space or efficient coupling with an optical fiber. Optical amplification of the received signal is also a possibility, as long as the signal level exceeds the amplifier's noise level. The pointing, acquisition, and tracking (PAT) assembly consists of a dichroic mirror to separate the transmit and receive beam, a two-axis fine-pointing mirror whose function is to remove the host platform vibrations, a two-axis point-ahead mirror to compensate for (point ahead or point behind) angle that arises from the cross velocity between the spacecraft and the target (e.g., Earth), and a focal plane array (FPA) consisting of as little as four pixels to hundreds of pixels. The fine-pointing (fast steering) mirror may require an update rate bandwidth of hundreds of hertz to several kilohertz, depending on the nature of the host platform vibration. It is also conceivable for the PAT's pixilated array to also serve as sensor for uplink data detection (Ortiz and Farr, 2009).

An alternative approach involves use of an isolation platform that combines passive and active isolators, as the interface between transceiver and the host platform (Chen et al., 2006). An advantage of this architecture is elimination of the fine-pointing mirror, since the base plate of the transceiver performs that function. Figure 7 illustrates such a configuration schematically.

The requirement for micro-radian level laser beam pointing necessitates optomechanical and thermomechanical integrity under the adverse conditions encountered in space. Maintaining optical isolation between the high-power transmit beam and the faint incoming beam, typically greater than 150 dB, is often a challenge. Even though transmit/receive signal isolation is more of a challenge for optical systems that combine both the transmitter and the (beacon and signal) receive aperture in one assembly, this configuration is preferred due to



Emerging Technologies, Free-Space Optical Communications, Figure 6 Block diagram example for a transceiver's optics assembly.



Emerging Technologies, Free-Space Optical Communications, Figure 7 Schematic of a laser communication transceiver utilizing isolation platform to simplify its architecture.

lower susceptibility for drift between channels (Chen and Lesh, 1994). A solar background, light-rejecting window at the telescope will minimize heat load from sunlight entering the telescope. Stringent requirements on cleanliness of the optics for flight use are an optical system class with cleanliness of 300 or better. This diminishes the scattered light that can find its way to any one of the sensors in the system (Spyak and Wolf, 1992). Redundancy of channels (if required for reliability reasons) may be implemented by polarization coupling for coupling two units and wavelength diversity for coupling additional units. Fiber coupling of all or some of the channels would simplify the optical assembly, while possibly encountering some additional coupling losses.

Pointing, acquisition, and tracking (PAT)

Fine pointing, acquisition, and tracking (PAT) is conceivably the most critical aspect of any lasercom link. This challenge stems from the narrow beamwidth of a communication beam, requiring the communication system to point its laser beam with sub-micro-radian (μrad) accuracy to a target tens to hundreds of millions of kilometers away. A combination of coarse and fine beam pointing is required to accomplish fine PAT. For planetary missions, a dedicated coarse-pointing mechanism is not required. The coarse-pointing (approximately, 0.1°), typically provided by the spacecraft to the planetary radio-frequency communications systems, is adequate for lasercom transceivers to acquire the Earth-based receiver within its acquisition field of view (Hemmati, 2006, 2009). Fine beam pointing and point ahead/behind is accomplished by way of internal mechanism(s) within the optomechanical assembly.

Prior to commencing a laser communications link between two remote transceivers, a number of knowledge-acquisition and handshaking steps need to be accomplished simultaneously:

- Knowing that the remote transceivers are within each other's field of view.
- Identifying each transceiver's position (e.g., attitude and locations). If range (and therefore roundtrip travel time) is short enough, the position knowledge may be communicated between transceivers.
- Utilizing the acquired position knowledge to derive the pointing vector to the remote transceiver.
- Deriving time-varying attitude variations and utilizing coarse-pointing and fine-pointing mechanisms to align the optomechanical assembly with the pointing vector.

Severely degraded link margin and system performance (signal fades at the receiver) will result from inaccurate laser beam pointing. The pointing loss allocation in the link control table is commonly divided into bias and jitter mispoint errors. Jitter represents high-frequency disturbances – normally root-sum squared to achieve an estimate of pointing accuracy – while bias represents the low bandwidth disturbances, normally added linearly.

The 3 sigma total pointing accuracy is calculated by adding $3 \times$ jitter to bias (Hemmati, 2007).

To account for the statistical pointing-induced fades (PIF) and mispoint angles, a pointing loss is typically allocated to the power link margin. For instance, the diffraction-limited angular beamwidth of a 1,000 nm laser traversed through a 30 cm-diameter telescope is approximately $3.3 \mu\text{rad}$. In this example, a 2 dB pointing loss allocation with a 1 % PIF requires a total (3 sigma) pointing accuracy of $2 \mu\text{rad}$. A promising new loss mitigation technique utilizes interleaver codes that spread the PIF over many code words of an error-correction code (Barron and Boroson, 2006).

Effective delivery of the spacecraft-emanated signal to a ground station requires the lasercom transceiver to track the receiving station such that the tracking jitter error is less than approximately 10 % of the transmit beamwidth (roughly $0.33 \mu\text{rad}$ in the above example) (Ortiz and Lee, 2006). For reference, from Mars, Earth angular extend is about $35 \mu\text{rad}$. Therefore, in the above example, the spaceborne laser beam needs to be pointed to approximately 1/18th of the Earth's diameter, a pointing requirement orders of magnitude more precise than required for RF system operating over the same range.

Spacecraft platform's angular microvibrations (with frequencies in the 0.1 Hz to hundreds of hertz range) stemming from scheduled and unscheduled guidance and control events (e.g., thruster firings and momentum dumping) constitute additional complexities for the already challenging sub-micro-radian level pointing requirement. Passive isolators and inertially referenced stabilized platforms can effectively mitigate the high (greater than several hundred hertz)-frequency vibrations (McMickell et al., 2007). A reliable beacon signal (e.g., laser beam, Sun-illuminated Earth or celestial references), in conjunction with a dedicated fine beam pointing control subsystem, is required to reject lower frequencies effectively (Hemmati, 2006, 2009).

The total laser beam pointing error for an optical transceiver is primarily comprised of inaccuracies in boresight calibration, the errors in position determination, and residual tracking errors not compensated for by the pointing subsystem control loop. A residual tracking error results from inadequate compensation of the host platform vibrations by the pointing control loop. The higher the bandwidth of the control loop, the better the platform jitter compensation and therefore the lower the residual laser beam pointing error. Data from inertial navigation subsystem (located in close proximity with the transceiver) can reduce the required image centroiding update rates from several kilohertz to hundreds of hertz or lower (Ortiz et al., 2001).

As indicated earlier, mounting the lasercom transceiver on a disturbance-free platform (DFP) effectively isolates the transceiver from the host platform, circumventing the need for fast-tracking capability (Scozzafava et al., 2007). Such architecture will minimize the amount of

line-of-sight jitter coupled into the optics assembly and allow the beacon signal analyzing focal plane array to be integrated over a much longer period of time, as well as enable the use of faint reference beacons as pointing references. With the ability for long integration times, laser beam pointing becomes considerably less sensitive to scintillation-induced fades of the uplink beacon, thereby relaxing the spacecraft pointing requirements. Moreover, a DFP can also provide coarse and fine-pointing adjustments, eliminating the need for additional high-bandwidth line-of-sight control (e.g., fast-steering mirror) within the system (Sannibale et al., 2009).

Optical channel (atmospherics)

The adverse effects on a laser beam traversing through the atmosphere may be put into two categories: (1) *attenuation effects*, caused by *scattering* and *absorption*, and (2) *refractive index fluctuation effects*, including spatial and temporal *scintillation*, *angle-of-arrival fluctuations*, large-scale *beam steering*, and *beam spreading*. The net effect of atmospheric attenuation and turbulence on communication link is signal fades, beam spreading at the telescope focal plane, wave front tip/tilt, background light due to sky radiance, signal scintillation, and other undesirable phenomena (Andrews and Phillips, 2005).

Attenuation reduces available signal power at the receiver and is a function of weather conditions, zenith angle, and the receiver site. Scattering is highly variable and depends on particle size in the atmosphere. If δ is the particle size and λ is laser's wavelength, different scattering cross sections (σ) can result depending on the relative size of δ and λ . A laser beam propagating through the atmosphere may experience Rayleigh scattering, which occurs when $\delta \ll \lambda$ (gases), Mie scattering when $\delta \sim \lambda$ (aerosols and fog), and geometric scattering when $\delta \gg \lambda$ (Andrews et al., 2001).

Scintillation is caused by refractive index variations in the atmosphere and results from spatial and temporal variations in light intensity caused by small-scale atmospheric turbulence. This effect creates multipath summation of signals at the receiver, resulting in phase differences for each beam with the net effect of variable arrival angles (Hemmati, 2006, 2009). Strong scintillation is difficult to characterize well, affects bit-error-rate performance, its influence increases with communications distance, and is more pronounced for horizontal links and when transmitting narrow beamwidth laser beams from the ground to space. Turbulence effects, scattering, and attenuation are all altitude dependent. At high altitudes, the effect is small for large telescopes, and turbulence along the path results in beam steering. Turbulence above the ground-based receiver results in wave front distortions in the form of a blur at the focal spot on the photodetector located behind the telescope. A significant portion of the atmospheric-induced optical link degradation can be allayed through proper site selection (Wojcik et al., 2005).

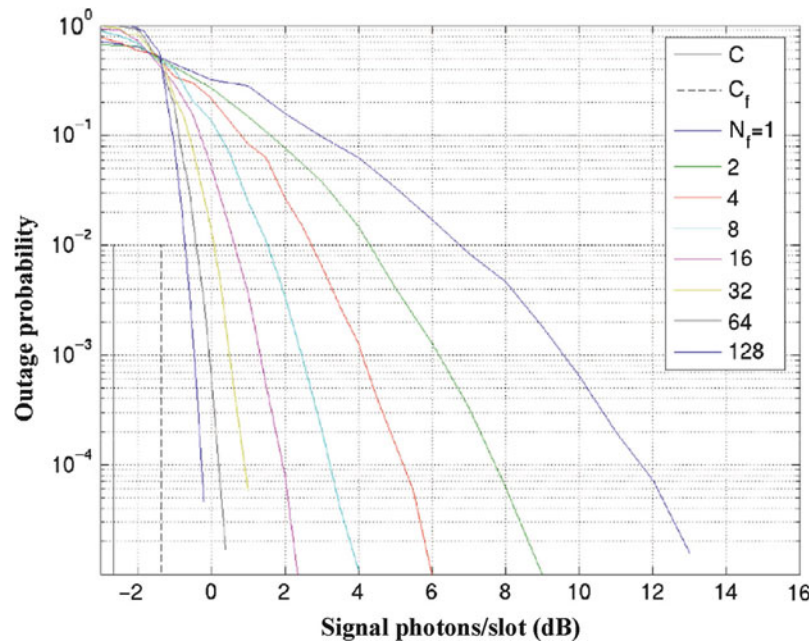
The atmospheric effects, singly or as an aggregate, can lead to dynamic attenuation of the signal strength. This phenomenon is referred to as signal fading. A fade is a transient condition that occurs when signal intensity at the receiver at any given instance is less than what is required to maintain the required link margin for a specified bit error rate (Andrews et al., 2001). Scintillation, spot dancing, and a time-varying speckle pattern at the detector are among major causes of fade events. The received signal fades away when a dark portion of the pattern overlaps the active region of the photodetector. Deep fades typically last 1–100 ms, depending on the strength of the turbulence effect for a particular link, and speed of the airflow across the beam driven by wind (Wojcik et al., 2005). In a high data rate optical link, significant amount of data could be lost during a fade period.

Fade mitigation techniques include use of large aperture diameters (greater than, approximately, 0.3 m), transmission of multiple beams that are mutually incoherent and physically separated (on the order of centimeters, use of long interleaver codes), implementation of adaptive optics technology, use of multimode lasers, and time or frequency diversity (Quaale et al., 2005; Moision and Biswas, 2007). As briefly explained below, each of these techniques suffers from implementation complexities or limitations.

For *downlink signal reception*, the *aperture averaging effect* (which becomes prominent when utilizing multimeter diameter ground telescopes) can mitigate scintillation effects on the beam and therefore decrease the fluctuations in total received power (Yuksel et al., 2005). Aperture averaging occurs when average size of the atmospheric turbulence cells is smaller than the diameter of the optical receiver. For *uplink beam transmission*, multibeam (4–16 beam) uplink can also mitigate some of the atmospheric turbulence effects on the beam (Wilson et al., 2007). For the multibeam fade mitigation to work, the laser beams have to be mutually incoherent and spatially separated enough to traverse uncorrelated atmospheric paths (Cameron et al., 2005).

Figure 8 shows that as much as 13 dB of the atmospheric fade depth can be mitigated via gains from implementation of long interleaver codes (Moision and Biswas, 2007).

Link availability is the fraction of time out of a year over which a laser communications link is operational. Link availability is highly dependent on the time of day and location of the telescope receiver site. Telescope site location options include space, platforms situated above clouds, and ground. Space-based receiving stations provide high link availability due to cloud-cover mitigation but suffer from limited aperture diameter and the lack of ability to upgrade and maintain the system after launch (Hurd et al., 2005; Biswas et al., 2006). High-altitude platforms avoid some of the drawbacks of space-based systems. But due to high winds that are typically encountered at those altitudes, these platforms are often challenged in providing lasercom-suitable



Emerging Technologies, Free-Space Optical Communications, Figure 8 Effect of interleaver in atmospheric fade mitigation is illustrated.

attitude-control levels. In the short term, ground-based receiver networks with atmospheric conditions limiting link availability provide an affordable solution. A recent study indicated that a network of five optical stations located globally will provide link availability of approximately 93 % (Link et al., 2004).

Optics: earth-based aperture

The Earth-based receiver collects the downlink signal photons and directs it to the photodetectors. Clearly, the larger the effective ground transceiver diameter, the greater the margin for the link. The excess link margin may be applied to shift hardware implementation burden away from the spacecraft (and onto the ground), to reduce flight system mass and power requirements, or to increase the amount of data delivered to Earth. The optical aperture may be a single telescope several meters in diameter or an array of distributed telescopes with meter-scale diameters. Each telescope may have a monolithic or a segmented primary mirror. Also, this transceiver may be located on the ground or at high altitudes above clouds, with clear advantages for each case.

The recent advancements in solid-state, very-low-noise, high-detection-efficiency photon-counting detectors (that operate near room temperature) have given credence to optical receiver architectures that utilize meter-class telescope diameters, each equipped on at least a 2×2 array of photon counters. Features of an array of telescope receivers are redundancy, graceful degradation, affordability scaled by the number of elements in the array, and potential for cost savings of a network of stations.

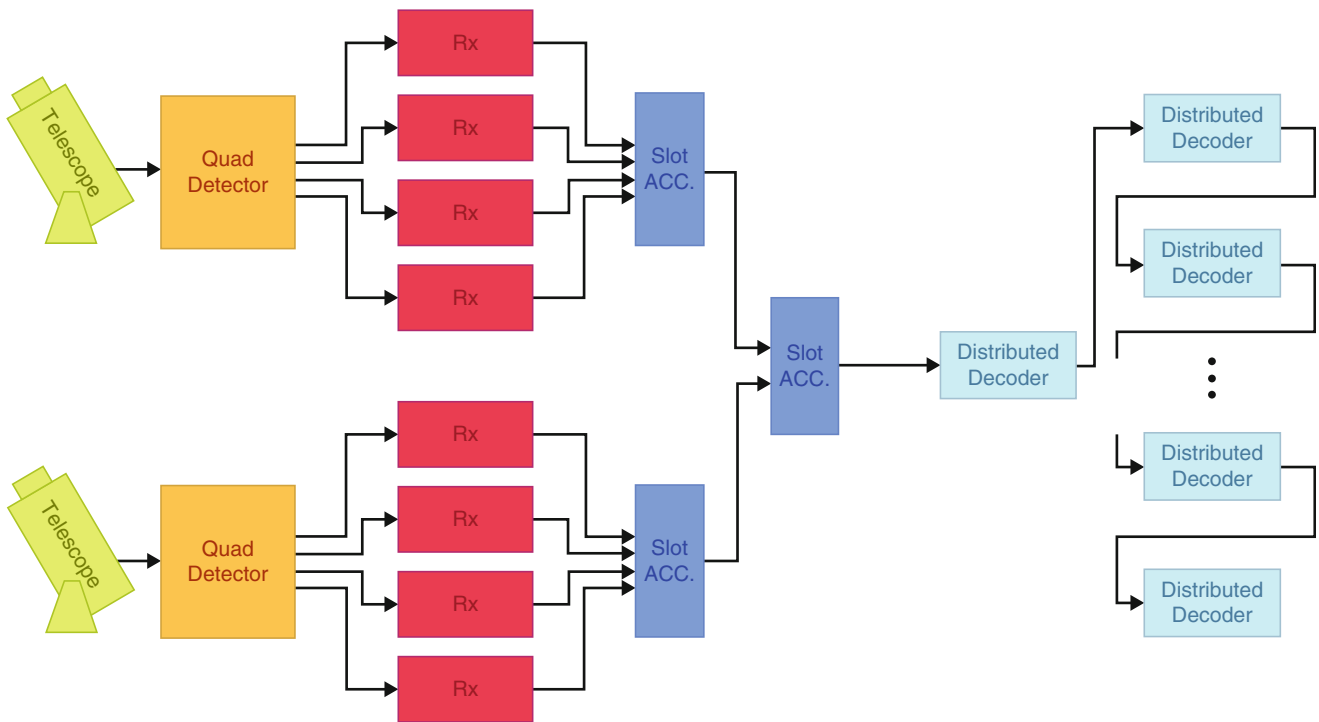
Figure 9 illustrates the schematic of a distributed and scalable ground receiver architecture.

The focal spot diameter of meter-scale diameter telescopes is set by the atmospheric turbulence-induced blurred conditions. The smallest achievable blur circle diameter for a downlink laser beam of wavelength λ that has traversed through an atmosphere with its seeing set by r_0 is set by an angle that is approximately $(r_0/\lambda)^2$ steradians (Roberts, 2005). Large-diameter membrane medium (10s of nm) band-pass optical filters (located at the entrance aperture of the telescope) and small-diameter narrow (sub-nm) band-pass filters, located just before the photodetectors(s), will combine to provide an essential level of background light rejection (Degnan, 1993).

Hybrid laser communications and precision laser ranging

An active laser-ranging transponder and laser communications transceiver have multiple-system-level requirements and subsystems in common. Steps necessary to operate a laser communications system in a transponder mode to infer millimeter-level range resolution are outlined below.

The interest in high-accuracy laser ranging to Mars is motivated by tests of relativistic gravitation, including the tests of the strong equivalence principle, Mars interior science (through exploitation of sensitivity to Mars precession, nutation, polar motion), and planetary science (through improvement of basic dynamical model parameters for the solar system). Centimeter-level resolution of Earth-Moon distance has been demonstrated via laser ranging to passive corner-cube retroreflectors deployed



Emerging Technologies, Free-Space Optical Communications, Figure 9 Scalable optical receiver followed by distributed decoder architecture.

on the lunar surface (Williams et al., 2009). Millimeter-level accuracy of lunar laser-ranging (LLR) data is expected in the near future (Turyshv and Williams, 2007; Murphy et al., 2007). However, due to the high space loss stemming from the $1/R^4$ energy transfer from passive corner cubes, passive laser ranging to planets is not possible at this time. Due to the benefit of $1/R^2$ energy transfer, active laser ranging to transponders located at the Moon or Mars will provide higher signal-to-noise ratios. The development of active laser-ranging techniques would extend the accuracies characteristic of passive laser tracking to interplanetary distances. Picosecond timing resolution time tagging of received and transmitted signal can now enable mm-class resolution ranging between the Earth and Mars.

A pulsed laser transponder system (based on time-of-flight measurements) is an efficient approach, requiring lower laser power than other available schemes and may be used in photon-counting mode without the need for phase referencing. A simple laser transponder might employ an “echo” technique, wherein the laser pulses from the remote unit are produced in response to detections of received pulses (Degnan, 1993). A more robust solution is paired one-way ranging, an asynchronous scheme in which both Earth and Moon/Mars stations transmit free-running pulse trains, time tagging all transmit and receive events (Degnan, 2002). Advantages of the asynchronous scheme include (1) noise immunity, the remote station does not transmit pulses in response to

numerous background photon detections; (2) stable laser performance, steady pulse generation results in thermally stable conditions; and (3) multiple Earth stations can range to the remote unit simultaneously without placing conflicting demands on the remote transmitter. With this scheme, clocks at each end of the link provide the baseline for time measurements, though only the Earth clock needs to keep absolute time. A frequency error in the remote clock is separable from velocity effects given multiple bidirectional measurements.

Despite the difference between beam pointing requirements for laser ranging and a lasercom transceiver, system-level requirements may be merged effectively to form a dual-functioning instrument serving both applications. Key differences between the two subsystems and methods to mitigate them are as follows:

1. Extensive transmit/receive isolation is typically required at each transceiver to maximize the link capacities of a bidirectional laser communications system. In a laser-ranging system, each transceiver must direct a portion of the transmitted light to its receiver in order to get an unbiased measurement of the transmit/receive time intervals.
2. Planetary communications links are typically designed for highly asymmetric performance, with a downlink data rate much higher than the uplink. A ranging system requires allocation of equal timing precision on each link.

3. The bandwidth and peak-to-average power ratio are chosen to maximize the data rate in a communications system. In a ranging system, a higher bandwidth always improves the precision at a given power, as does a higher peak-to-average power ratio in the presence of background light. Thus, a ranging system will typically require shorter laser pulses, higher pulse energies at lower repetition rates, and higher detector bandwidths (lower jitter).

Techniques to mitigate each of the above possible conflicts have been identified and are being implemented in the laboratory (Hemmati et al., 2009).

Summary

Technology of free-space laser communications at the component, subsystem, and system levels has advanced enough to successfully validate the lasercom technology at the maximum Mars range with data rates approaching 500 Mb/s – nearly two orders of magnitude beyond the state-of-the-art radio-frequency communications systems. There is opportunity for significant additional link margin improvement. When realized, the additional link margin may be used to reduce mass, power consumption, size, and complexity of the flight transceiver. The enhanced data rate capability will, for example, enable streaming high-quality video from other planets to the Earth. The “light science” enabled by lasercom (akin to radio science) remains a largely unexplored field.

Key remaining engineering challenges include circumventing attenuation due to dense clouds/fog, communicating while the receiver field of view is within a 1° of the Sun, and addressing the inadequacy of ground station infrastructure. While future Earth-based optical receivers may be located on airborne or Earth-orbiting platforms to circumvent atmospheric issues, technology and operational demonstrations for the first decade or so of deep-space laser communications will likely rely on lower risk and less-expensive ground-based receivers. Ground station diversity can adequately mitigate cloud cover. Large-diameter membrane filters located at the entrance aperture of ground-based telescopes, or telescopes designed to circumvent Sun interference, will aid with communications at small Sun angles. The advent of affordable, large-diameter, non-diffraction-limited ground telescopes should make infrastructure development realizable.

Acknowledgments

The author is indebted to the members of the Optical Communications Group and associates at JPL whose work is reported here, particularly W. Farr, A. Biswas, J. Kovalik, G. Ortiz, K. Wilson, S. Piazzolla, W. T. Roberts, M. Wright, K. Birnbaum, B. Moision, J. Hamkins, K. Quirk, J. Gin, S. Townes, F. Pollara, P. Estabrook, D. Antsos, and F. Davarian.

Bibliography

- Andrews, L. C., and Phillips, R. L., 2005. *Laser Beam Propagation through Random Media*, 2nd edn. Bellingham: SPIE Press.
- Andrews, L. C., Phillips, R. L., and Hopen, C. Y., 2001. *Laser Beam Scintillation with Applications*. Bellingham: SPIE Press.
- Barron, R. J., and Boroson, D. M., 2006. Analysis of capacity and probability of outage for free-space optical channels with fading due to pointing and tracking error. *Proceedings of SPIE*, **6105**, 61050B.
- Biswas, A., Wright, M. W., Kovalik, J., and Piazzolla, S., 2005. Uplink beacon laser for Mars laser communication demonstration (MLCD). *Proceedings of SPIE*, **5712**, 931–100.
- Biswas, A., Khatri, F., and Boroson, D., 2006. Near-Sun free-space optical communications from space. In *IEEE Aerospace Conference* paper# 1611.
- Boroson, D., 2007. A survey of technology-driven capacity limits for free-space laser communications. *Proceedings of SPIE*, **6709**, 670918.
- Braun, A., Rudd, J. V., Cheng, H., Mourou, G., Kopf, D., Jung, I. D., Weingarten, K. J., and Keller, U., 1995. Characterization of short-pulse oscillators by means of a high-dynamic-range autocorrelation measurement. *Optics Letters*, **20**, 1889.
- Bromage, J., Winzer, P. J., Nelson, L. E., Mermelstein, M. D., Horn, C., and Headley, C. H., 2003. Amplified spontaneous emission in pulsed-pumped Raman amplifiers. *IEEE Photonics Technology Letters*, **15**, 667–669.
- Cameron, A., Dios, F., Rodriguez, A., Rubio, J. A., Reyes, M., and Alonso, A., 2005. Modeling of power fluctuations induced by refractive turbulence in a multiple-beam ground-to-satellite optical uplink. *Proceedings of SPIE*, **5892**, 196–205.
- Caplan, D. O., 2007. Laser communication transmitter and receiver design. *Journal of Optical Fiber Communications*, **4**, 225–362.
- Charpylyv, A. R., and Tkach, R. W., 1993. What is the actual capacity of single-mode fibers in amplified lightwave systems. *IEEE Photonics Technology Letters*, **5**, 666–668.
- Chen, C.-C., and Lesh, J. R., 1994. Overview of the optical communications demonstrator. *Proceedings of SPIE*, **2123**, 85–94.
- Chen, C.-C., Hemmati, H., Biswas, A., Ortiz, G., Farr, W., and Pedreiro, N., 2006. Simplified lasercom system architecture using a disturbance-free platform. *Proceedings of SPIE*, **6105**, 610505.
- Degnan, J. J., 1993. Millimeter accuracy satellite laser ranging: a review. In Smith, D. E., and Turcotte, D. L. (eds.), *Contributions of Space Geodesy to Geodynamics: Technology, Geodyn. Ser.*, vol. 25, pp. 133–162, AGU, Washington, DC.
- Degnan, J. J., 2002. Asynchronous laser transponders for precise interplanetary ranging and time transfer. *Journal of Geodynamics*, **34**, 551.
- Deng, Y., Koch, M. W., Lu, F., Wicks, G. W., and Knox, W. H., 2004. Colliding-pulse passive harmonic mode-locking in a femtosecond Yb-doped fiber laser with a semiconductor saturable absorber. *Optics Express*, **12**, 3872.
- Dolinar, S., Divsalar, D., Hamkins, H., and Pollara, F., 2000. Capacity of pulse position modulation (PPM) on gaussian and Wenn channels. *The Interplanetary Network Directorate Progress Report 42–142*, Jet Propulsion Laboratory, Pasadena, California, pp. 1–31. <http://ipnpr.jpl.nasa.gov/progress-report/>.
- Dupriez, P., Piper, A., Malinowski, A., Sahu, J. K., Ibsen, M., Thomsen, B. C., Jeong, Y., Hickey, L. M. B., Zervas, M. N., Nilsson, J., and Richardson, D. J., 2006. High average power, high repetition rate, picosecond pulsed fiber master oscillator power amplifier source seeded by a gain-switched Laser diode at 1060 nm. *IEEE Photonics Technology Letters*, **18**, 1013–1015.
- Farr, W. H., 2009. Negative avalanche feedback detectors for photon-counting optical communications. *Proceedings of SPIE*, **7199**, 23.

- Gin, J. W., Nguyen, D. H., and Quirk, K. J., 2007. High data rate receiver for optical PPM communications. In *NASA Earth Science and Technology Conference Proceedings*, College Park, NSTC-07-0140.
- Hamkins, J., 2008. Pulse position modulation. In Bidgoli, H. (ed.), *The Handbook of Computer Networks*. Hoboken: Wiley.
- Hamkins, J., and Moision, B., 2004. Selection of modulation and codes for deep space optical communications. *Proceedings of SPIE*, **5338**, 123–130. San Jose, CA (USA).
- Hemmati, H. (ed.), 2006. *Deep Space Optical Communications*. Hoboken: Wiley.
- Hemmati, H., 2007. Interplanetary laser communications. *Optics & Photonics News*, **18**, 22–27. Washington D. C, (USA).
- Hemmati, H. (ed.), 2009. *Near-Earth laser Communications*. Boca Raton: CRC Press. Chapter 10.
- Hemmati, H., and Lesh, J. R., 1998. A combined laser communication and imager for micro-spacecraft (ACLAIM). *Proceedings of SPIE*, **3266**, 165. San Jose, CA (USA).
- Hemmati, H., Wright, M. W., Biswas, A., and Esproles, C., 2000. High-efficiency pulsed laser transmitters for deep-space communication. *Proceedings of SPIE*, **3932**, 188–195. San Jose, CA (USA).
- Hemmati, H., Biswas, A., and Boroson, D. M., 2007. Prospects for improvement of interplanetary laser communication data rates by 30dB. *Proceedings of IEEE*, **95**, 2082–2092. 10.1109/JPROC.2007.905057
- Hemmati, H., Birnbaum, K. M., Farr, W. H., Turyshv, S., and Biswas, A., 2009. Combined laser-communications and laser-ranging transponder for Moon and Mars. *Proceedings of SPIE*, **7199**, 71990N-1 to N-12.
- Hurd, W. J., MacNeal, B. E., Ortiz, G. G., Cheng, E. S., Moe, R. V., Walker, J. Z., Fairbrother, D., Dennis, M., Eegholm, B., and Kasunic, K. J., 2005. Exo-atmospheric telescopes for deep-space optical communications. In *IEEE Aerospace Conference*, paper # 1557.
- Kachelmyer, A. L., and Boroson, D. M., 2007. Efficiency penalty of photon-counting with timing jitter. *Proceedings of SPIE*, **6709**, 670906.
- Kerman, A. J., Dauler, E. A., Keicher, W. E., Yang, J. K., Berhhren, K. K., Gol'tsman, G., and Voronov, B., 2006. Kinetic-inductance-limited reset time of superconducting nanowire photon counters. *Applied Physics Letters*, **88**, 111116–111119.
- Kiasaleh, K., 1998. Turbo-coded optical PPM communications. *Journal of Lightwave Technology*, **16**, 18–26.
- Link, R., Alliss, R., and Craddock, M. E., 2004. Mitigating the impact of clouds on optical communications. *Proceedings of SPIE*, **5338**, 223.
- Lipes, R. G., 1980. Pulse-position-modulation coding as near-optimum utilization of photon counting channel with bandwidth and power constraints. In *DSN Reporting Program*. Vol. 42–56, pp. 108–113.
- McMickell, M. B., Kreider, T., Hansen, E., Davis, T., and Gonzalez, M., 2007. Optical payload isolation using the miniature vibration isolation system. *Proceedings of SPIE*, **6527**, 652703.
- Moision, B., 2008a. Communication limits due to photon detector jitter. *IEEE Photonics Technology Letters*, **20**, 715–717.
- Moision, B., 2008. Universal decoder for variable duty-cycle optical communications., In *JPL's IPN Progress Report*, pp. 42–173.
- Moision, B., and Biswas, A., 2007. Memory requirements to mitigate fading loss on an optical channel. In *JPL's IPN Progress Report*, pp. 42–170.
- Moision, B., and Farr, W. H., 2008. Communication losses due to photon-detector jitter. *IEEE Photonics Letters*, **20**, 715–717.
- Moision, B., and Hamkins, J., 2002. Constrained coding for the deep-space optical channel. *Proceedings of SPIE*, **4635**, 127–137. San Jose, CA (USA).
- Moision, B., and Hamkins, J., 2003. Deep-space optical communications downlink budget: modulation & coding. In *IPN Progress Report 42–154*. pp. 1–28.
- Moision, B., Hamkins, J., and Cheng, M., 2006. Design of a coded modulation for deep space optical communications. In *Information Theory and Applications*. San Diego: University of California, February 6–10, 2006.
- Murphy, T. W., Michelsen, E. L., Orin, A. E., Adelberger, E. G., Hoyle, C. D., Swanson, H. E., Stubbs, C. W., and Battat, J. E., 2007. APOLLO: a new push in lunar laser ranging. In *Proceedings of the International Workshop From Quantum to Cosmos: Fundamental Physics Research in Space*. May 22–24, 2006, Warrenton, International Journal of Modern Physics D, **16** (12a), 2127.
- Ortiz, G. G., and Farr, W. H., Charles, J. R., Roberts, W. T., Sannibale, V., Gin, J., Saharaspude, A., and Garkanian, V., 2009. “Canonical deep space optical communications transceiver”, *Proc. SPIE 7199*, Free-Space Laser Communication Technologies XXI, 71990K (February 24,2009); doi:10.1117/12.812410; <http://dx.doi.org/10.1117/12.812410>.
- Ortiz, G. G., and Lee, S., 2006. Chapter 5.3, deep-space acquisition, tracking and pointing. In Hemmati, H. (ed.), *Deep Space Optical Communications*. Hoboken: Wiley.
- Ortiz, G. G., Le, S., and Alexander, J. W., 2001. Sub-microradian pointing for deep-space optical telecommunications network. In *19th AIAA International Communications Satellite Systems Conference*. Toulouse.
- Quaale, R., Hindman, B., Engberg, B., and Collier, P., 2005. Mitigating environmental effects on free-space laser communications. In *IEEE Aerospace Conference*. Montana.
- Quirk, K. J., and Gin, J. W., 2006. Optical PPM combining loss for photon counting receivers. In *Military Communications Conference, IEEE (MILCOM)*. pp. 1–4.
- Roberts, W. T., 2005. Optical membrane technology for deep space optical communications filters. In *IEEE Aerospace Conference Proceedings*. Big Sky, 1991–2000.
- Robinson, B., Caplan, D. O., Stevens, M. L., Barron, R. J., Dauler, E. A., and Hamilton, S. A., 2005. 1.5-photons/bit photon-counting optical communications using Geiger-mode avalanche photodiodes. In *IEEE LEOS Summer Topical Meetings*.
- Sannibale, V., Ortiz, G. G., Farr, W. H., 2009. A sub-Hertz vibration isolation platform (LFVIP) for a deep space optical communication transceiver. *Proceedings of SPIE*, **7199**, 7199–17.
- Scozzafava, J. J., Boroson, D. M., Burnside, J. W., Glynn, M. L., DeVoe, C. E., DeFranzo, C. M., and Doyle, K. B., 2007. Design of a very small inertially stabilized optical space terminal. *Proceedings of SPIE*, **6709**, 670905. San Jose, CA (USA).
- Smith, D. E., Zuber, M. T., Sun, X., Neumann, G. A., Cavanaugh, J. F., McGarry, J. F., and Zagwodzki, T. W., 2006. Two-way laser link over interplanetary distance. *Science*, **311**(5757), 53.
- Smutny, B., Kaempfer, H., Muehlnikel, G., Sterr, U., Wandernoth, B., Heine, F., Hildebrand, U., Dallma, D., Reinhardt, M., Freier, A., Lange, R., Boehmer, K., Feldhaus, T., Mueller, J., Weichert, A., Greulich, P., Seel, S., Meyer, R., and Czichy, R., 2009. 5.6 Gbps optical intersatellite communication link. *Proceedings of SPIE*, **7199**, 719906. San Jose, CA (USA).
- Spellmeyer, N. W., Caplan, D. O., and Stevens, M. L., 2005. Design for a 5-watt PPM transmitter for the mars laser communications demonstration. In *Digest of the LEOS Summer Topical Meetings*. Paper TuA4.1, pp. 51–52.
- Spyak, P. R., and Wolf, W. L., 1992. Scatter from particulate-contaminated mirrors. Part 1: theory and experiment for polystyrene spheres and $\lambda = 0.632 \mu\text{m}$. *Optical Engineering*, **31**, 1746.
- Stern, J. A., and Farr, W. H., 2007. Fabrication and characterization of superconducting NbN nanowire single photon detector. *IEEE Transactions on Applied Superconductivity*, **17**, 306–309.

- Tan, Y., Guo, J., Ai, Y., Liu, W., and Fei, Y., 2008. A coded modulation scheme for deep-space optical communications. *IEEE Photonics Technical Letters*, **20**, 372–374.
- Tolker Nielsen, T., and Oppenhauser, G., 2002. In orbit test result of an operational optical intersatellite link between ARTEMIS and SPOT4, SILEX. *Proceedings of SPIE*, **4635**, 1. San Jose, CA (USA).
- Turyshv, S. G., and Williams, J. G., 2007. Space-based tests of gravity with laser ranging. *International Journal of Modern Physics*, **D16**, 2165–2179.
- Williams, J. G., Turyshv, S. G., and Boggs, D. H., 2009. Lunar laser ranging tests of the equivalence principle with the earth and moon. *International Journal of Modern Physics*, **D18**, 1129–1175.
- Wilson, K. E., Lesh, J. R., and Yan, T.-Y., 1993. GOPEX: a laser uplink to the Galileo spacecraft on its way to Jupiter. *Proceedings of SPIE*, **1866**, 138. San Jose, CA (USA).
- Wilson, K., Lesh, J. R., Araki, K., and Arimoto, Y., 1997. Overview of ground-to-orbit lasercom demonstration. *Proceedings of SPIE*, **2990**, 23–30.
- Wilson, K., Troy, M., Srinivasan, M., Platt, B., Vilnrotter, V., Wright, M., Garkanian, V., Hemmati, H., 2003. Daytime adaptive optics for deep space optical communications. In *Conference Proceedings International Space Conference of Pacific-basin Societies (ISCOPS)*, Tokyo.
- Wilson, K., Kovalik, J., Biswas, A., and Roberts, W. T., 2007. Development laser beam transmission strategies for ground-to-space optical communications. *Proceedings of SPIE*, **6551**, 65510B. San Jose, CA (USA).
- Winzer, P. J., Kalmar, A., and Leeb, W. R., 1999. Role of amplified spontaneous emission in optical free-space communication links with optical amplification – impact on isolation and data transmission; utilization for pointing acquisition and tracking. *Proceedings of SPIE*, **3615**, 34–141. San Jose, CA (USA).
- Wojcik, G. S., Szymczak, H. L., Alliss, R. J., Link, R. P., Craddock, M. E., and Mason, M. L., 2005. Deep-space to ground laser communications in a cloudy world. *Proceedings of SPIE*, **5892**, 589203. San Jose, CA (USA).
- Wyner, A. D., 1988. Capacity and error exponent for the direct detection photon channel – part I. *IEEE Transactions on Information Theory*, **34**, 1449–1461.
- Wysocki, P., Wood, T., Grant, A., Holcomb, D., Chang, K., Santo, M., Braun, L., and Johnson, G., 2006. High reliability 49 dB gain, 13 W PM fiber amplifier at 1550 nm with 30 dB PER and record efficiency. In *Optical Fiber Communication Conference*. Long Beach, paper PDP17.
- Yuksel, H., Milner, S., and Davis, C. C., 2005. Aperture averaging for optimizing receiver design and system performance on free-space optical communication links. *Journal of Optical Networking*, **4**, 462–475.

EMERGING TECHNOLOGIES, LIDAR

David M. Tratt
The Aerospace Corporation, Los Angeles, CA, USA

Synonyms

Ladar; Laser radar; Optical radar

Definitions

Lidar. Light Detection And Ranging.

Ladar. LAsER Detection And Ranging.

Introduction

The term *lidar* (i.e., the optical analogue of radar) embraces a broad collection of active optical probing techniques for remotely determining the physical and chemical characteristics of diffuse or condensed matter targets, both natural and otherwise. Although lidar methods have been practiced since before the dawn of the laser era, their classification here as an emerging technology reflects the fact that they have only in relatively recent years begun to see systematic exploitation from space-based vantage points (most especially with the advent of robust, high-performance laser systems).

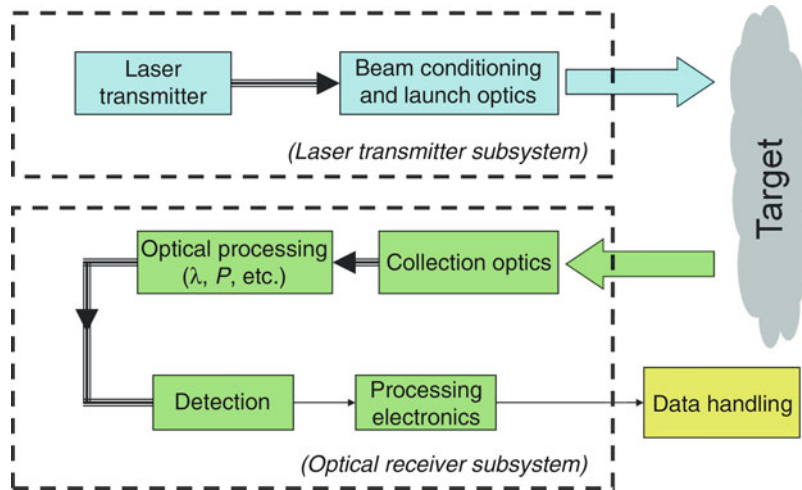
Some remarks on terminology are in order as a guide to those unfamiliar with this field. Although the terms *lidar* and *ladar* are essentially interchangeable (except for exceedingly rare instances in which an incoherent source is used and which would therefore exclude consideration of *ladar*), different application communities have preferentially adopted one over the other. Hence, the atmospheric science community uses the form *lidar*, while those working in areas that involve remote characterization of hard targets have adopted *laser radar* or *ladar* (McManamon, 2012). It is also possible to find instances where *ladar* and *lidar* appear in upper case, reflecting their origin as acronyms. This is becoming less common now that these terms (like *radar* before them) have entered general parlance, but the hybrid form **LIDAR** is nevertheless commonplace albeit used exclusively by the airborne laser terrain mapping community.

A number of textbooks are available that treat lidar systems and methodology from a variety of standpoints. Measures (1992) provides a thorough grounding in the basic theory and practice of laser remote sensing, while Fujii and Fukuchi (2005), Weitkamp and Walther (2005), and Kovalev and Eichinger (2004) collectively offer a comprehensive overview of the wide array of measurement applications that have benefitted from lidar. Grant et al. (1997) provide an anthology of seminal lidar papers from the earliest days, when arc lamps were the brightest sources available, up to the mid-1990s.

Lidar basics

A canonical lidar system comprises the set of subsystems depicted in Figure 1. Depending on the intended application, the laser transmitter may require stringent control of wavelength, spectral purity, polarization, pulse duration, or some combination of all of these properties. By corollary, the receiver system may also be required to analyze one or all of these same properties of the return radiation.

Following the detection stage, processing electronics convert the signals to a range-gated (i.e., time-of-flight registered) data stream which is then logged for further processing to extract the desired observable quantity. This is accomplished through an understanding of the often complex interaction phenomenology of the laser radiation with the target and frequently leads to a suboptimally constrained retrieval process which must be augmented



Emerging Technologies, Lidar, Figure 1 Functional schematic of a generic lidar system.

by ancillary measurements, modeled parameters, or a combination of both. The optical power $P(R)$ returned to the lidar detection subsystem by a scattering volume located at range R is given by the standard basic lidar equation:

$$P(R) = \left[E\eta O(R) \frac{cA}{2R^2} \right] [T(R)T'(R)\beta(\lambda, R)], \quad (1)$$

where E is the pulse energy emitted into the atmosphere (outward transmission T , return transmission T'), A is the receiver collecting area, and β is the backscatter from the range element of interest at the specified emission wavelength, λ . η is the overall optical efficiency of the receiver subsystem and $O(R)$ is the range-dependent transmit/receive overlap function. (In general, the optical efficiency represents the optical transmission. However, for a *coherent* detection lidar system, this parameter will also include a component due to heterodyne mixing efficiency (Zhao et al., 1990).) The distinction made between outgoing and return atmospheric transmission recognizes that the incident radiation in some cases will be inelastically scattered, as explained below. Terms grouped together in the first set of square parentheses in Equation 1 are lidar system-dependent parameters, while those contained in the second set of square parentheses characterize the atmospheric response. This expression also assumes single scattering. Many lidar systems are subject to multiple scattering to a greater or lesser extent, and this will complicate the data inversion procedure. Eloranta (1998) provides a discussion of these issues and techniques for their resolution.

While pulsed operation is implicitly assumed in Equation 1, other forms of range gating have used modulation of continuous wave (cw) laser sources, such as pseudorandom noise encoding (Takeuchi et al., 1986), in conjunction with synchronous detection of the return to extract the time-dependent signal behavior.

Most, if not all, lidars are operated in the backscattering or quasi-backscattering configuration in which the transmitter and receiver are collocated and form part of an integral instrument. When the transmit and receive optical axes are coaxial, the lidar is said to be *monostatic*, whereas if they are offset, the lidar is described as *bistatic*. Note that this nomenclature differs from that applied in the radar field in that a bistatic radar system comprises transmitter and receiver elements that are geographically separated, whereas a monostatic radar implies collocated transmitter and receiver systems that may or may not be coaxial. Multistatic lidar configurations comprising two or more receiver stations have been considered for certain specialized applications (e.g., Olofson et al., 2008), but thus far, none have been implemented beyond the experimental stage.

An additional distinction that is frequently used to classify lidar systems is the mode of photon detection employed. In a *direct (incoherent) detection* lidar, the received photons are converted directly to electrical impulses by the detector with the loss of all phase information, and this is the detection method found in the overwhelming majority of lidar systems. In a *heterodyne (coherent) detection* lidar, the received photons are mixed with a reference oscillator signal and detection is performed on the intermediate frequency. Such a configuration is technologically more complex than a direct detection system, and also makes considerably more stringent demands on laser frequency stability and beam quality, but preserves phase knowledge and is essential if the ultimate goal is to directly retrieve motion-induced Doppler signatures, such as for atmospheric wind measurement (Huffaker and Hardesty, 1996) or object tracking (Osche and Young, 1996). A few coherent detection lidars have been described that are non-Doppler capable (e.g., Grant et al., 1987; Menzies and Tratt, 1994; Gibert et al., 2006). In these cases the additional system

complexity is tolerated because of the much greater noise rejection that is afforded by coherent detection, which can be as much as an order of magnitude above that of a comparable direct detection system (Menzies, 1976).

Inelastic backscatter lidar

Inelastic backscatter lidars make use of wavelength conversion properties of the target to infer details of its physical and chemical makeup. The chief downshifting processes that are commonly employed in the lidar field are Raman and fluorescence scattering, both of which require excitation in the UV-visible spectral range.

Although the weak cross sections for Raman scattering tend to limit its usefulness for long-range remote sensing purposes, this disadvantage is offset by the high specificity that it affords (because the observed wavelength shift is diagnostic of the molecule being probed). Hence, many users tolerate the need for a high-power transmitter, large collection aperture, and long accumulation intervals in order to gain the benefit of unambiguous measurements, which is clearly of central importance when such systems are used to validate satellite data products (Wessel et al., 2000; Tratt et al., 2005). Raman lidar has met with significant success in the ground-based remote sensing of atmospheric constituents and structure (Turner and Whiteman, 2006), atmospheric dynamics (Koch et al., 1991), and hazardous substances (Sharma et al., 2005) and has also seen limited airborne application (Heaps and Burris, 1996).

Fluorescence lidars are generally used for the investigation of chemical and biological targets. For example, laser-induced fluorescence has been successfully used in the remote assessment of plant health (Saito et al., 2000), airborne monitoring of phytoplankton populations and ocean productivity (Hoge et al., 2005), and the diagnosis of biodeterioration in building fabrics (Weibring et al., 2001). These techniques rely on measurement of the 690 nm chlorophyll- α fluorescence signature (sometimes in association with other features) and an understanding of the way it is modified by environmental disturbances. Fluorescence lidar has also been proposed for monitoring of marine oil spills (Karpicz et al., 2006) and the identification of airborne chemical agents or pathogens (Buteau et al., 2007).

So-called resonance fluorescence techniques, as the name implies, resonantly excite a target atomic species in order to directly probe the atmospheric properties at the residence altitude of the species in question, which are typically metal ions entrained in the middle and upper atmosphere (80–100 km altitude) that are derived from, and continually replenished by, meteoritic deposition. Hence, there is a long heritage of using sodium lidars to track atmospheric dynamics in these difficult-to-access regions (Gardner, 1989) as well as to retrieve upper atmospheric temperatures through measurement of the Doppler-broadened atomic lineshape, which can then be related directly to the ambient temperature. Other atomic metal species that have been studied for similar purposes

are potassium (Papen et al., 1995) and iron (Kane and Gardner, 1993). More recently, iron Boltzmann lidars have been used to infer atmospheric temperature by measuring the spectral content of radiation emitted by resonantly excited mesospheric iron layer atoms (Gelbwachs, 1994; Raizada and Tepley, 2002). An extensive treatment of this field has been given by Chu and Papen (2005).

An intriguing new class of inelastic backscatter lidar is the so-called femtosecond white-light lidar. This employs an ultrashort-pulse transmitter operating in the near-IR at around 800 nm whose high intensity gives rise to nonlinear self-phase modulation that results in broadband upconverted continuum emission across the UV-IR spectral range (Kasparian et al., 2005). The critical factor in such systems is that the broadband continuum is emitted predominantly in the backward direction and is thus a particularly strong and effective backscattering mechanism (Yu et al., 2001). Applications for the femtosecond lidar center on investigations requiring multiple wavelength channels to constrain multivariate retrievals, such as high-resolution atmospheric absorption spectrum measurement and aerosol composition and microphysics determination (Kasparian et al., 2003).

Elastic backscatter lidar

The majority of lidar systems in use can be broadly classed as elastic backscatter systems. Elastic backscattering processes conserve the centroid wavelength and consist of Rayleigh (or molecular) scattering from air molecules and particle scattering from suspended particulate material (aerosols and clouds). (The term Mie scattering is frequently used to implicitly denote particle scattering; however, the Mie descriptor strictly speaking only applies to scattering from *spherical* particles, which in general account for only a small percentage of the total atmospheric aerosol population.) Rayleigh scattering magnitude is inversely proportional to the fourth power of wavelength, which in practice limits the usefulness of Rayleigh lidar approaches to wavelengths not exceeding the visible spectrum. Rayleigh lidar systems tend to be used for profiling fundamental atmospheric properties such as temperature and density (Chanin and Hauchecorne, 1984).

By far the commonest application of lidar is concerned with cloud and aerosol (i.e., particulate) detection and characterization. In its simplest form, this can entail measurements at a single wavelength to infer the broad characteristics of scattering structure in atmospherically entrained dust and cloud features. However, in the absence of ancillary data, such minimally constrained measurements can only acquire estimates of attenuated backscatter. To properly retrieve aerosol optical properties and physical parameters requires a multiwavelength lidar approach and calls for backscatter coefficients at a minimum of three wavelengths and aerosol extinction coefficients for at least two wavelengths (e.g., Müller et al., 1998; Althausen et al., 2000). Aerosol extinction

has typically been obtained from collocated, contemporaneous column optical depth measurements (using sunphotometers) or from auxiliary Raman lidar channels, which measure the Raman backscatter from a well-mixed atmospheric constituent (usually nitrogen) and observe its departure from expected values to ascribe the aerosol extinction (Ansmann et al., 1992). However, in recent years there has been increasing interest in the high spectral resolution lidar (HSRL) technique because of its ability to separate and directly measure the aerosol and molecular scattering components with a single, self-contained multiwavelength instrument that avoids recourse to ancillary measurements (Eloranta, 2005; Hair et al., 2008).

In addition to the wealth of ground-based and airborne cloud/aerosol lidar systems that are today operated around the world, some functioning as components of autonomous, long-term monitoring networks, three multiwavelength systems have flown in Earth orbit, one on the Space Shuttle (Winker et al., 1996) and two on dedicated free-flying platforms (Spinhirne et al., 2005; Winker et al., 2007). There has even been a cloud/aerosol lidar operated on the surface of Mars (Whiteway et al., 2008).

Polarimetric lidar

The shape of cloud/aerosol particles can also modify the polarization state of scattered radiation. Polarization lidars exploit this phenomenon by separately measuring the unaltered (p) and depolarized (s) components of the return radiation, for which a range-dependent *backscatter depolarization ratio*, $\delta(R)$, can then be defined:

$$\delta(R) = \frac{P_s(R)}{P_p(R)}. \quad (2)$$

(Note, however, that Gimmestad (2008) has offered a recasting of the traditional polarization lidar formulation that is more descriptive of the microphysical processes involved and which may in time supersede this notation.)

The magnitude of the depolarization ratio can be used to deduce the degree of nonsphericity of the particles. The relative polarimetric properties of various aerosol and cirrus particles show sufficient diversity that polarization lidar can provide for some level of discrimination between the different classes of such particles (Sassen, 1991; Sassen, 2007). However, care must be taken when inferring scattering or attenuation coefficients from lidar measurements of cirrus, as the preferential horizontal orientation of cirrus platelets has been shown to dramatically enhance lidar returns through the addition of a specular reflection component to the backscatter signal (Platt et al., 1978). In order to separate the intrinsic cirrus scattering properties from this effect, it is therefore advisable to offset the lidar viewing angle by $\sim 3^\circ$ from the vertical.

Differential absorption lidar

The differential absorption lidar (DIAL) evolved as a means for identifying and quantifying gaseous

atmospheric constituents. The technique entails transmitting two different wavelengths, one tuned to a characteristic spectral feature of the target species and the other offset by a small but sufficient amount that it is unaffected by the target gas and so acts as an atmospheric transmissivity reference. The usual practice is to contrive the on- and off-resonant wavelengths to be as narrowly separated as possible (in order to eliminate dispersive effects from other atmospheric constituents that could confuse the measurement), so that the following close approximation for the number density $N(R)$ of absorbing molecules can be applied:

$$N(R) = \frac{1}{2|\sigma|} \frac{d \ln [P_{on}(R)/P_{off}(R)]}{dR}, \quad (3)$$

in which σ is the differential absorption cross section between the on- and off-resonant wavelengths.

A large number of DIAL systems are in operation worldwide to measure a wide variety of atmospheric trace gases from the ground and from the air. Because of their climatic impacts, water vapor and ozone profiling are two common DIAL measurement objectives (Browell et al., 2004), while a plethora of systems operating from the UV to the thermal-IR are in use to track the transport and dispersion of plumes containing common pollutants such as sulfur dioxide (Weibring et al., 1998), ammonia (Zhao et al., 2002), hydrocarbons (Murdock et al., 2008), and many other compounds.

The ongoing debate over the role of carbon dioxide in climate change has driven a resurgence in interest concerning DIAL techniques for monitoring this atmospheric constituent. The requirement is for mixing ratio measurement globally throughout the lowermost levels of the troposphere with a precision of $\sim 0.5\%$, which is considerably more challenging than the typical $\sim 20\%$ DIAL precision specification. Conventional DIAL is one of two approaches that are under evaluation for this application (Koch et al., 2004; Gibert et al., 2006), the other being integrated path differential absorption (IPDA), which entails cw or quasi-cw illumination of the Earth's surface at on- and off-resonant probe wavelengths (Menzies and Tratt, 2003). The retrieved differential absorption return can then be transformed into a path-averaged column content (Abshire et al., 2010; Spiers et al., 2011). Detailed trade studies will ascertain the optimum wavelength combination and detection configuration for acquiring high-precision gas mixing ratios by this technique.

Doppler lidar

Doppler lidars are used to measure target motion, either for atmospheric wind measurement (Baker et al., 1995) or object tracking (Osche and Young, 1996), and can be divided into two distinct classes. *Coherent detection Doppler lidars* directly measure the Doppler shift imposed on the optical return by the atmospheric target by mixing the return signal with a well-defined frequency reference

laser, termed a *local oscillator*. The heterodyne mixing efficiency in such systems is a sensitive function of frequency purity, wave front quality, and optical aberration factors, any one of which can be the source of significant performance degradation if not adequately managed (Zhao et al., 1990; Chambers, 1997). For hard target coherent Doppler lidar systems, where backscatter is high, these considerations are less of an issue, but for atmospheric applications, where performance is contingent on aerosol and cloud backscatter, one is effectively driven to operating wavelengths in the IR spectral region.

By contrast, *direct detection Doppler lidars* operate in the “photon bucket” mode and thus do not mandate the wave front quality requirements of coherent systems. In a direct detection system, the Doppler shift is inferred indirectly by measuring the differential transmission of the return signal through a high-resolution filter mechanism whose dispersion characteristics must be known to high accuracy and very well controlled. Freed from the IR wavelength constraint to which coherent systems are subject, direct detection Doppler lidars are thus designed to take advantage of the omnipresence of molecules in the atmosphere by operating in Rayleigh backscatter mode (usually in the eye-safe UV). Direct detection systems are less photon efficient than their coherent counterparts, and their retrieval accuracy can be compromised by the individual and collective stability of optical components within the receiver subsystem, but are preferred for operation against atmospheres that have little or no aerosol content. The combining of both Doppler lidar approaches in a *hybrid* instrument has been suggested as a means to make best use of the advantages of both types of system (Emmitt, 2004).

There is a significant body of experience with both surface-based (Huffaker and Hardesty, 1996; Grund et al., 2001; Henderson and Hannon, 2005) and airborne (Bilbro et al., 1986; Rothermel et al., 1998; Werner et al., 2001; Hannon et al., 1999) coherent Doppler wind lidars. In addition, there have been several ground-based direct detection Doppler wind lidars (von Zahn et al., 2000; Gentry et al., 2000; Irgang et al., 2002) and also a few airborne systems (Dehring et al., 2006; Durand et al., 2006). All require some means of scanning the transceiver in order to obtain the multiple perspective line-of-sight retrievals that are necessary to resolve 2D or 3D wind vector fields.

The European Space Agency (ESA) plans to deploy the first space-based Doppler lidar. The Atmospheric Dynamics Mission (ADM-Aeolus) will carry a direct detection Doppler wind lidar and an integral high spectral resolution lidar (HSRL) with two receiver channels for analysis of aerosol/cloud backscatter and molecular backscatter, respectively (Stoffelen et al., 2005). This instrument will address only a single off-nadir line-of-sight perspective but will nevertheless represent the first on-orbit demonstration of Doppler wind lidar techniques.

It is worth noting here that the first lidar wind measurements from space were made by a non-Doppler elastic backscatter system using optical scatterometry of the

ocean surface (Menzies et al., 1998). These measurements were necessarily indirect, scalar in character, and restricted to sea surfaces only, but more recent measurements indicate that this approach can provide significant new information on ocean winds and associated air–sea interaction processes (Hu et al., 2008).

Laser ranging and geodetic imaging

Laser ranging to hard targets was among the earliest of lidar applications and has seen by far the most commercial exploitation. Apart from the proliferation of laser-augmented rangefinding binoculars and surveying instruments, the technique is now in high demand for airborne high-density terrain mapping (Slatton et al., 2007) and shallow-water hydrographic surveying (Banic and Cunningham, 1998), for which a significant commercial services capability has arisen. Laser altimetry was also the first lidar application to be implemented in space (Sjogren and Wollenhaupt, 1973; Bufton, 1989; Garvin et al., 1998), and global-scale geodetic mapping by laser altimetry can in many respects be regarded as commonplace, having now been demonstrated with moderate horizontal resolution (10–100 m) on Earth (Schutz et al., 2005), Mars (Smith et al., 2001), Mercury (Zuber et al., 2008), the Moon (Smith et al., 1997; Smith et al., 2010), and the irregular asteroids Eros (Cheng et al., 2002) and Itokawa (Mukai et al., 2007). Most of these were first-return time-of-flight instruments, the exception being the Geoscience Laser Altimeter System (GLAS), which was designed to transcribe the entire return waveform (Abshire et al., 2005). Full waveform recovery makes it possible for laser altimetry to resolve volumetric vegetation structure and sub-canopy topography by taking advantage of the optical porosity of vegetation cover (Blair et al., 1999; Lefsky and McHale, 2008).

Some measurement scenarios call for a level of areal density of individual ranging measurements that pushes horizontal resolution toward the meter range. For these applications, limited pulse energy and the need for rapid scanning mean that “photon starvation” becomes a significant issue (Abrams and Tratt, 2005). To address this concern, single-photon counting detector arrays with integrated readout circuitry have been developed to acquire 3D imagery of laser flood-illuminated scenes in what is often termed a *flash lidar* scheme (Albota et al., 2002). In addition, work is currently in process to extend active image formation techniques from the established synthetic aperture radar domain into the optical spectral region (Beck et al., 2005). This technique faces nontrivial technical challenges but is nevertheless being pursued for tactical applications (Ricklin and Tomlinson, 2005) and has also been proposed for planetary remote sensing (Karr, 2003).

High-density laser ranging is also used in space-to-space scenarios where the uncompromising demands of autonomous rendezvous and proximity operations rule out less precise instrumental approaches (Nimelman et al., 2006).

Summary

Active optical remote sensing (lidar) techniques have provided demonstrable enhancements in spatiotemporal resolution and reduced dependence on the diurnal cycle in comparison to other remote sensing methods used for Earth and planetary science applications. While the breadth and diversity of lidar methodology attests to its fundamental versatility, there are certain measurements that are enabled uniquely by active optical sensors (e.g., aerosol and cloud vertical structure, global-scale tropospheric wind measurement, 3D vegetation structure, high-resolution sub-canopy topography mapping). The intrinsic benefits of lidar now drive a steadily increasing number of fielded systems on the ground, in the air, and in space. While not exhaustive, this entry describes the basic characteristics and associated applications of the commonest types of lidar likely to be encountered and also provides an introduction to some promising emerging concepts that have yet to fully mature into accepted field-capable systems.

Bibliography

- Abrams, M. C., and Tratt, D. M., 2005. Progress in laser sources for lidar applications: laser sources for 3D imaging remote sensing. *Proceedings of SPIE*, **5653**, 241, doi:10.1117/12.580323.
- Abshire, J. B., Sun, X., Riris, H., Sirota, J. M., McGarry, J. F., Palm, S., Yi, D., and Liiva, P., 2005. Geoscience laser altimeter system (GLAS) on the ICESat mission: on-orbit measurement performance. *Geophysical Research Letters*, **32**, L21S02, doi:10.1029/2005GL024028.
- Abshire, J. B., Riris, H., Allan, G. R., Weaver, C. J., Mao, J., Sun, X., Hasselbrack, W. E., Kawa, S. R., and Biraud, S., 2010. Pulsed airborne lidar measurements of atmospheric CO₂ column absorption. *Tellus*, **62B**, 770, doi:10.1111/j.1600-0889.2010.00502.x.
- Albota, M. A., Heinrichs, R. M., Kocher, D. G., Fouche, D. G., Player, B. E., O'Brien, M. E., Aull, B. F., Zayhowski, J. J., Mooney, J., Willard, B. C., and Carlson, R. R., 2002. Three-dimensional imaging laser radar with a photon-counting avalanche photodiode array and microchip laser. *Applied Optics*, **41**, 7671, doi:10.1364/AO.41.007671.
- Althausen, D., Müller, D., Ansmann, A., Wandinger, U., Hube, H., Clauer, E., and Zörner, S., 2000. Scanning 6-wavelength 11-channel aerosol lidar. *Journal of Atmospheric and Oceanic Technology*, **17**, 1469, doi:10.1175/1520-0426(2000)017<1469:SWCAL>2.0.CO;2.
- Ansmann, A., Wandinger, U., Riebesell, M., Weitkamp, C., and Michaelis, W., 1992. Independent measurement of extinction and backscatter profiles in cirrus clouds by using a combined Raman elastic-backscatter lidar. *Applied Optics*, **31**, 7113, doi:10.1364/AO.31.007113.
- Baker, W. E., Emmitt, G. D., Robertson, F., Atlas, R. M., Molinari, J. E., Bowdle, D. A., Paegle, J., Hardesty, R. M., Menzies, R. T., Krishnamurti, T. N., Brown, R. A., Post, M. J., Anderson, J. R., Lorenc, A. C., and McElroy, J., 1995. Lidar-measured winds from space: a key component for weather and climate prediction. *Bulletin of the American Meteorological Society*, **76**, 869, doi:10.1175/1520-0477(1995)076<0869:LMWFS>2.0.CO;2.
- Banic, J. R., and Cunningham, A. G., 1998. Airborne laser bathymetry: a tool for the next millennium. *EEZ Technology*, **3**, 75. http://www.jalbtcx.org/downloads/Publications/32Banic_Cunningham_98.pdf.
- Beck, S. M., Buck, J. R., Buell, W. F., Dickinson, R. P., Kozlowski, D. A., Marechal, N. J., and Wright, T. J., 2005. Synthetic-aperture imaging laser radar: laboratory demonstration and signal processing. *Applied Optics*, **44**, 7621, doi:10.1364/AO.44.007621.
- Bilbro, J. W., DiMarzio, C. A., Fitzjarrald, D. E., Johnson, S. C., and Jones, W. D., 1986. Airborne Doppler lidar measurements. *Applied Optics*, **25**, 3952, doi:10.1364/AO.25.003952.
- Blair, J. B., Rabine, D. L., and Hofton, M. A., 1999. The laser vegetation imaging sensor: a medium-altitude, digitization-only, airborne laser altimeter for mapping vegetation and topography. *ISPRS Journal of Photogrammetry and Remote Sensing*, **54**, 115, doi:10.1016/S0924-2716(99)00002-7.
- Browell, E. V., Grant, W. B., and Ismail, S., 2004. Environmental measurements: laser detection of atmospheric trace gases. In Guenther, R. D., Steel, D. G., and Bayvel, L. (eds.), *Encyclopedia of Modern Optics*. Amsterdam: Elsevier, pp. 403–416.
- Buften, J. L., 1989. Laser altimetry measurements from aircraft and spacecraft. *Proceedings of the IEEE*, **77**, 463, doi:10.1109/5.24131.
- Buteau, S., Simard, J.-R., Lahaie, P., Roy, G., Mathieu, P., Déry, B., Ho, J., and McFee, J., 2007. Bioaerosol standoff monitoring using intensified range-gated laser-induced fluorescence spectroscopy. In Kim, Y. J., and Platt, U. (eds.), *Advanced Environmental Monitoring*. Dordrecht: Springer, pp. 203–216, doi:10.1007/978-1-4020-6364-0_16.
- Chambers, D., 1997. Modeling of heterodyne efficiency for coherent laser radar in the presence of aberrations. *Optics Express*, **1**, 60, doi:10.1364/OE.1.000060.
- Chanin, M. L., and Hauchecorne, A., 1984. Lidar studies of temperature and density using rayleigh scattering. In *International Council of Scientific Unions Middle Atmosphere Handbook*, Vol. 13, Urbana, Illinois, USA: Scientific Committee on Solar Terrestrial Physics, International Council of Scientific Unions, p. 87.
- Cheng, A. F., Barnouin-Jha, O., Prockter, L., Zuber, M. T., Neumann, G., Smith, D. E., Garvin, J., Robinson, M., Veverka, J., and Thomas, P., 2002. Small-scale topography of 433 Eros from laser altimetry and imaging. *Icarus*, **155**, 51, doi:10.1006/icar.2001.6750.
- Chu, X., and Papen, G. C., 2005. Resonance fluorescence lidar for measurements of the middle and upper atmosphere. In Fujii, T., and Fukuchi, T. (eds.), *Laser Remote Sensing*. Boca Raton: Taylor & Francis, pp. 179–432, doi:10.1201/9781420030754.ch5.
- Dehring, M. T., Tchoryk, P., and Wang, J., 2006. High altitude balloon-based wind LIDAR demonstration: from near space to space. *Proceedings of SPIE*, **6220**, 62200P, doi:10.1117/12.669262.
- Durand, Y., Chinal, E., Endemann, M., Meynart, R., Reitebuch, O., and Treichel, R., 2006. ALADIN airborne demonstrator: a Doppler wind lidar to prepare ESA's ADM-Aeolus explorer mission. *Proceedings of SPIE*, **6296**, 62961D, doi:10.1117/12.680958.
- Eloranta, E. W., 1998. Practical model for the calculation of multiply scattered lidar returns. *Applied Optics*, **37**, 2464, doi:10.1364/AO.37.002464.
- Eloranta, E. W., 2005. High spectral resolution lidar. In Weitkamp, C., and Walther, H. (eds.), *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*. Berlin: Springer, pp. 143–163, doi:10.1007/0-387-25101-4_5.
- Emmitt, G. D., 2004. Combining direct and coherent detection for Doppler wind lidar. *Proceedings of SPIE*, **5575**, 31, doi:10.1117/12.576539.
- Fujii, T., and Fukuchi, T. (eds.), 2005. *Laser Remote Sensing*. Boca Raton: Taylor & Francis. <http://www.crcnetbase.com/isbn/9781420030754>.

- Gardner, C. S., 1989. Sodium resonance fluorescence lidar applications in atmospheric science and astronomy. *Proceedings of the IEEE*, **77**, 408, doi:10.1109/5.24127.
- Garvin, J., Bufton, J., Blair, J. B., Harding, D., Luthcke, S., Frawley, J., and Rowlands, D., 1998. Observations of the Earth's topography from the shuttle laser altimeter (SLA): laser-pulse echo-recovery measurements of terrestrial surfaces. *Physics and Chemistry of the Earth*, **23**, 1053, doi:10.1016/S0079-1946(98)00145-1.
- Gelbwachs, J. A., 1994. Iron Boltzmann factor lidar: proposed new remote sensing technique for mesospheric temperature. *Applied Optics*, **33**, 7151, doi:10.1364/AO.33.007151.
- Gentry, B. M., Chen, H., and Li, S. X., 2000. Wind measurements with 355-nm molecular Doppler lidar. *Optics Letters*, **25**, 1231, doi:10.1364/OL.25.001231.
- Gibert, F., Flamant, P. H., Bruneau, D., and Loth, C., 2006. Two-micrometer heterodyne differential absorption lidar measurements of the atmospheric CO₂ mixing ratio in the boundary layer. *Applied Optics*, **45**, 4448, doi:10.1364/AO.45.004448.
- Gimmetstad, G. G., 2008. Reexamination of depolarization in lidar measurements. *Applied Optics*, **47**, 3795, doi:10.1364/AO.47.003795.
- Grant, W. B., Margolis, J. S., Brothers, A. M., and Tratt, D. M., 1987. CO₂ DIAL measurements of water vapor. *Applied Optics*, **26**, 3033, doi:10.1364/AO.26.003033.
- Grant, W. B., Browell, E. V., Menzies, R. T., Sassen, K., She, C.-Y., and Thompson, B. J. (eds.), 1997. *Selected Papers on Laser Applications in Remote Sensing*. Bellingham: SPIE.
- Grund, C. J., Banta, R. M., George, J. L., Howell, J. N., Post, M. J., Richter, R. A., and Weickmann, A. M., 2001. High-resolution Doppler lidar for boundary layer and cloud research. *Journal of Atmospheric and Oceanic Technology*, **18**, 376, doi:10.1175/1520-0426(2001)018<0376:HRDLFB>2.0.CO;2.
- Hair, J. W., Hostetler, C. A., Cook, A. L., Harper, D. B., Ferrare, R. A., Mack, T. L., Welch, W., Izquierdo, L. R., and Hovis, F. E., 2008. Airborne high spectral resolution lidar for profiling aerosol optical properties. *Applied Optics*, **47**, 6734, doi:10.1364/AO.47.006734.
- Hannon, S. M., Bagley, H. R., and Bogue, R. K., 1999. Airborne Doppler lidar turbulence detection: ACLAIM flight test results. *Proceedings of SPIE*, **3707**, 234, doi:10.1117/12.351378.
- Heaps, W. S., and Burris, J., 1996. Airborne Raman lidar. *Applied Optics*, **35**, 7128, doi:10.1364/AO.35.007128.
- Henderson, S. W., and Hannon, S. M., 2005. Advanced coherent lidar system for wind measurements. *Proceedings of SPIE*, **5887**, 58870I, doi:10.1117/12.620318.
- Hoge, F. E., Lyon, P. E., Wright, C. W., Swift, R. N., and Yungel, J. K., 2005. Chlorophyll biomass in the global oceans: airborne lidar retrieval using fluorescence of both chlorophyll and chromophoric dissolved organic matter. *Applied Optics*, **44**, 2857, doi:10.1364/AO.44.002857.
- Hu, Y., Stamnes, K., Vaughan, M., Pelon, J., Weimer, C., Wu, D., Cisewski, M., Sun, W., Yang, P., Lin, B., Omar, A., Flittner, D., Hostetler, C., Trepte, C., Winker, D., Gibson, G., and Santa-Maria, M., 2008. Sea surface wind speed estimation from space-based lidar measurements. *Atmospheric Chemistry and Physics*, **8**, 3593, doi:10.5194/acp-8-3593-2008.
- Huffaker, R. M., and Hardesty, R. M., 1996. Remote sensing of atmospheric wind velocities using solid-state and CO₂ coherent laser systems. *Proceedings of the IEEE*, **84**, 181, doi:10.1109/5.482228.
- Huffaker, R. M., Jelalian, A., and Thomson, J. A. L., 1970. Laser-Doppler system for detection of aircraft trailing vortices. *Proceedings of the IEEE*, **58**, 322, doi:10.1109/PROC.1970.7636.
- Irgang, T. D., Hays, P. B., and Skinner, W. R., 2002. Two-channel direct-detection Doppler lidar employing a charge-coupled device as a detector. *Applied Optics*, **41**, 1145, doi:10.1364/AO.41.001145.
- Kane, T. J., and Gardner, C. S., 1993. Structure and seasonal variability of the nighttime mesospheric Fe layer at midlatitudes. *Journal of Geophysical Research*, **98**, 16875, doi:10.1029/93JD01225.
- Karpicz, R., Dementjev, A., Kuprionis, Z., Pakalnis, S., Westphal, R., Reuter, R., and Gulbinas, V., 2006. Oil spill fluorosensing lidar for inclined onshore or shipboard operation. *Applied Optics*, **45**, 6620, doi:10.1364/AO.45.006620.
- Karr, T. J., 2003. Synthetic aperture lidar for planetary sensing. *Proceedings of SPIE*, **5151**, 44, doi:10.1117/12.505723.
- Kasparian, J., Rodriguez, M., Méjean, G., Yu, J., Salmon, E., Wille, H., Bourayou, R., Frey, S., André, Y. B., Mysyrowicz, A., Sauerbrey, R., Wolf, J.-P., and Wöste, L., 2003. White light filaments for atmospheric analysis. *Science*, **301**, 61, doi:10.1126/science.1085020.
- Kasparian, J., Bourayou, R., Frey, S., Luderer, J. C., Méjean, G., Rodriguez, M., Salmon, E., Wille, H., Yu, J., Wolf, J.-P., and Wöste, L., 2005. Femtosecond white-light lidar. In Fujii, T., and Fukuchi, T. (eds.), *Laser Remote Sensing*. Boca Raton: Taylor & Francis, pp. 37–62, doi:10.1201/9781420030754.ch2.
- Koch, S. E., Dorian, P. B., Ferrare, R., Melfi, S. H., Skillman, W. C., and Whiteman, D., 1991. Structure of an internal bore and dissipating gravity current as revealed by Raman lidar. *Monthly Weather Review*, **119**, 857, doi:10.1175/1520-0493(1991)119<0857:SOAIBA>2.0.CO;2.
- Koch, G. J., Barnes, B. W., Petros, M., Beyon, J. Y., Amzajerdian, F., Yu, J., Davis, R. E., Ismail, S., Vay, S., Kavaya, M. J., and Singh, U. N., 2004. Coherent differential absorption lidar measurements of CO₂. *Applied Optics*, **43**, 5092, doi:10.1364/AO.43.005092.
- Kovalev, V. A., and Eichinger, W. E., 2004. *Elastic Lidar: Theory, Practice, and Analysis Methods*. Hoboken: Wiley, doi:10.1002/0471643173.
- Lefsky, M., and McHale, M. R., 2008. Volume estimates of trees with complex architecture from terrestrial laser scanning. *Journal of Applied Remote Sensing*, **2**, 023521, doi:10.1117/1.2939008.
- McManamon, P. F., 2012. Review of lidar: a historic, yet emerging, sensor technology with rich phenomenology. *Optical Engineering*, **51**, 060901, doi:10.1117/1.OE.51.6.060901.
- Measures, R. M., 1992. *Laser Remote Sensing: Fundamentals and Applications*. New York: Krieger.
- Menzies, R. T., 1976. Laser heterodyne detection techniques. In Hinkley, E. D. (ed.), *Laser Monitoring of the Atmosphere*. Berlin/Heidelberg: Springer, pp. 297–353, doi:10.1007/3-540-07743-X_21.
- Menzies, R. T., and Tratt, D. M., 1994. Airborne CO₂ coherent lidar for measurements of atmospheric aerosol and cloud backscatter. *Applied Optics*, **33**, 5698, doi:10.1364/AO.33.005698.
- Menzies, R. T., and Tratt, D. M., 2003. Differential laser absorption spectrometry for global profiling of tropospheric carbon dioxide: selection of optimum sounding frequencies for high-precision measurements. *Applied Optics*, **42**, 6569, doi:10.1364/AO.42.006569.
- Menzies, R. T., Tratt, D. M., and Hunt, W. H., 1998. Lidar In-space technology experiment measurements of sea surface directional reflectance and the link to surface wind speed. *Applied Optics*, **37**, 5550, doi:10.1364/AO.37.005550.
- Mukai, T., Abe, S., Hirata, N., Nakamura, R., Barnouin-Jha, O., Cheng, A., Mizuno, T., Hiraoka, K., Honda, T., and Demura, H., 2007. An overview of the LIDAR observations of asteroid 25143 Itokawa. *Advances in Space Research*, **40**, 187, doi:10.1016/j.asr.2007.04.075.
- Müller, D., Wandinger, U., Althausen, D., Mattis, I., and Ansmann, A., 1998. Retrieval of physical particle properties from lidar

- observations of extinction and backscatter at multiple wavelengths. *Applied Optics*, **37**, 2260, doi:10.1364/AO.37.002260.
- Murdock, D. G., Stearns, S. V., Lines, R. T., Lenz, D., Brown, D. M., and Philbrick, C. R., 2008. Applications of real-world gas detection: airborne natural gas emission lidar (ANGEL) system. *Journal of Applied Remote Sensing*, **2**, 023518, doi:10.1117/1.2937078.
- Nimelman, M., Tripp, J., Allen, A., Hiemstra, D. M., and McDonald, S. A., 2006. Spaceborne scanning lidar system (SSLS) upgrade path. *Proceedings of SPIE*, **6201**, 62011V, doi:10.1117/12.666187.
- Olofson, K. F. G., Witt, G., and Pettersson, J. B. C., 2008. Bistatic lidar measurements of clouds in the nordic arctic region. *Applied Optics*, **47**, 4777, doi:10.1364/AO.47.004777.
- Osche, G. R., and Young, D. S., 1996. Imaging laser radar in the near and far infrared. *Proceedings of the IEEE*, **84**, 103, doi:10.1109/5.482225.
- Papen, G. C., Gardner, C. S., and Pfenninger, W. M., 1995. Analysis of a potassium lidar system for upper-atmospheric wind-temperature measurements. *Applied Optics*, **34**, 6950, doi:10.1364/AO.34.006950.
- Platt, C. M. R., Abshire, N. L., and McNice, G. T., 1978. Some microphysical properties of an ice cloud from lidar observations of horizontally oriented crystals. *Journal of Applied Meteorology*, **17**, 1220, doi:10.1175/1520-0450(1978)017<1220:SMPOAI>2.0.CO;2.
- Raizada, S., and Tepley, C. A., 2002. Iron Boltzmann lidar temperature and density observations from Arecibo – an initial comparison with other techniques. *Geophysical Research Letters*, **29**, 1560, doi:10.1029/2001GL014535.
- Ricklin, J. C., and Tomlinson, P. G., 2005. Active imaging at DARPA. *Proceedings of SPIE*, **5895**, 589505, doi:10.1117/12.622572.
- Rothermel, J., Cutten, D. R., Hardesty, R. M., Menzies, R. T., Howell, J. N., Johnson, S. C., Tratt, D. M., Olivier, L. D., and Banta, R. M., 1998. The multi-center airborne coherent atmospheric wind sensor. *Bulletin of the American Meteorological Society*, **79**, 581, doi:10.1175/1520-0477(1998)079<0581:TMCACA>2.0.CO;2.
- Saito, Y., Saito, R., Kawahara, T. D., Nomura, A., and Takeda, S., 2000. Development and performance characteristics of laser-induced fluorescence imaging lidar for forestry applications. *Forest Ecology and Management*, **128**, 129, doi:10.1016/S0378-1127(99)00280-7.
- Sassen, K., 1991. The polarization lidar technique for cloud research: a review and current assessment. *Bulletin of the American Meteorological Society*, **72**, 1848, doi:10.1175/1520-0477(1991)072<1848:TPLTFC>2.0.CO;2.
- Sassen, K., 2007. Identifying atmospheric aerosols with polarization lidar. In Kim, Y. J., and Platt, U. (eds.), *Advanced Environmental Monitoring*. Berlin: Springer, pp. 136–142, doi:10.1007/978-1-4020-6364-0_10.
- Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., and DiMarzio, J. P., 2005. Overview of the ICESat mission. *Geophysical Research Letters*, **32**, L21S01, doi:10.1029/2005GL024009.
- Sharma, S. K., Misra, A. K., and Sharma, B., 2005. Portable remote Raman system for monitoring hydrocarbon, gas hydrates and explosives in the environment. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **61**, 2404, doi:10.1016/j.saa.2005.02.020.
- Sjogren, W. L., and Wollenhaupt, W. R., 1973. Lunar shape via the apollo laser altimeter. *Science*, **179**, 275, doi:10.1126/science.179.4070.275.
- Slatton, K. C., Carter, W. E., Shrestha, R. L., and Dietrich, W., 2007. Airborne laser swath mapping: achieving the resolution and accuracy required for geosurficial research. *Geophysical Research Letters*, **34**, L23S10, doi: 10.1029/2007GL031939.
- Smith, D. E., Zuber, M. T., Neumann, G. A., and Lemoine, F. G., 1997. Topography of the Moon from the clementine lidar. *Journal of Geophysical Research*, **102**, 1591, doi:10.1029/96JE02940.
- Smith, D. E., Zuber, M. T., Frey, H. V., Garvin, J. B., Head, J. W., Muhleman, D. O., Pettengill, G. H., Phillips, R. J., Solomon, S. C., Zwally, H. J., Banerdt, W. B., Duxbury, T. C., Golombek, M. P., Lemoine, F. G., Neumann, G. A., Rowlands, D. D., Aharonson, O., Ford, P. G., Ivanov, A. B., Johnson, C. L., McGovern, P. J., Abshire, J. B., Afzal, R. S., and Sun, X. L., 2001. Mars orbiter laser altimeter: experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research*, **106**, 23689, doi:10.1029/2000JE001364.
- Smith, D. E., Zuber, M. T., Neumann, G. A., Lemoine, F. G., Mazarico, E., Torrence, M. H., McGarry, J. F., Rowlands, D. D., Head III, J. W., Duxbury, T. H., Aharonson, O., Lucey, P. G., Robinson, M. S., Barnouin, O. S., Cavanaugh, J. F., Sun, X., Liiva, P., Mao, D.-D., Smith, J. C., and Bartels, A. E., 2010. Initial observations from the lunar orbiter laser altimeter (LOLA). *Geophysical Research Letters*, **37**, L18204, doi:10.1029/2010GL043751.
- Spiers, G. D., Menzies, R. T., Jacob, J., Christensen, L. E., Phillips, M. W., Choi, Y., and Browell, E. V., 2011. Atmospheric CO₂ measurements with a 2 μ m airborne laser absorption spectrometer employing coherent detection. *Applied Optics*, **50**, 2098, doi:10.1364/AO.50.002098.
- Spinhrne, J. D., Palm, S. P., Hart, W. D., Hlavka, D. L., and Welton, E. J., 2005. Cloud and aerosol measurements from GLAS: overview and initial results. *Geophysical Research Letters*, **32**, L22S03, doi:10.1029/2005GL023507.
- Stoffelen, A., Pailleux, J., Källénc, E., Vaughan, J. M., Isaksen, I., Flamant, P., Wergen, W., Andersson, E., Schyberg, H., Culoma, A., Meynard, R., Endemann, M., and Ingmann, P., 2005. The atmospheric dynamics mission for global wind field measurement. *Bulletin of the American Meteorological Society*, **86**, 73, doi:10.1175/BAMS-86-1-73.
- Takeuchi, N., Baba, H., Sakurai, K., and Ueno, T., 1986. Diode-laser random-modulation cw lidar. *Applied Optics*, **25**, 63, doi:10.1364/AO.25.000063.
- Tratt, D. M., Whiteman, D. N., Demoz, B. B., Farley, R. W., and Wessel, J. E., 2005. Active Raman sounding of the Earth's water vapor field. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **61**, 2335, doi:10.1016/J.SAA.2005.02.032.
- Turner, D. D., and Whiteman, D. N., 2006. Remote Raman spectroscopy. Profiling water vapor and aerosols in the troposphere using Raman lidars. In Chalmers, J. M., and Griffiths, P. R. (eds.), *Handbook of Vibrational Spectroscopy*. Chichester: Wiley, doi:10.1002/0470027320.s6803.
- Von Zahn, U., von Cossart, G., Fiedler, J., Fricke, K. H., Nelke, G., Baumgarten, G., Rees, D., Hauchecorne, A., and Adolfsen, K., 2000. The ALOMAR Rayleigh/Mie/Raman lidar: objectives, configuration, and performance. *Annales Geophysicae*, **18**, 815, doi:10.1007/s00585-000-0815-2.
- Weibring, P., Edner, H., Svanberg, S., Cecchi, G., Pantani, L., Ferrara, R., and Caltabiano, T., 1998. Monitoring of volcanic sulphur dioxide emissions using differential absorption lidar (DIAL), differential optical absorption spectroscopy (DOAS) and correlation spectroscopy (COSPEC). *Applied Physics B*, **67**, 419, doi:10.1007/s003400050525.
- Weibring, P., Johansson, T., Edner, H., Svanberg, S., Sundner, B., Raimondi, V., Cecchi, G., and Pantini, L., 2001. Fluorescence lidar imaging of historical monuments. *Applied Optics*, **40**, 6111, doi:10.1364/AO.40.006111.
- Weitekamp, C., and Walther, H. (eds.), 2005. *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere*. Berlin: Springer, doi:10.1007/b106786.

- Werner, C., Flamant, P. H., Reitebuch, O., Köpp, F., Streicher, J., Rahm, S., Nagel, E., Klier, M., and Herrmann, H., 2001. Wind infrared Doppler lidar instrument. *Optical Engineering*, **40**, 115, doi:10.1117/1.1335530.
- Wessel, J., Beck, S. M., Chan, Y. C., Farley, R. W., and Gelbwachs, J. A., 2000. Raman lidar calibration for the DMSP SSM/T-2 microwave water vapor sensor. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 141, doi:10.1109/36.823908.
- Whiteway, J., Daly, M., Carswell, A., Duck, T., Dickinson, C., Komguem, L., and Cook, C., 2008. Lidar on the phoenix mission to Mars. *Journal of Geophysical Research*, **113**, E00A08, doi:10.1029/2007JE003002.
- Winker, D. M., Couch, R. H., and McCormick, M. P., 1996. An overview of LITE: NASA's Lidar in-space technology experiment. *Proceedings of the IEEE*, **84**, 164, doi:10.1109/5.482227.
- Winker, D. M., Hunt, W. H., and McGill, M. J., 2007. Initial performance assessment of CALIOP. *Geophysical Research Letters*, **34**, L19803, doi:10.1029/2007GL030135.
- Yu, J., Mondelain, D., Ange, G., Volk, R., Niedermeier, S., Wolf, J. P., Kasparian, J., and Sauerbrey, R., 2001. Backward supercontinuum emission from a filament generated by ultrashort laser pulses in air. *Optics Letters*, **26**, 533, doi:10.1364/OL.26.000533.
- Zhao, Y., Post, M. J., and Hardesty, R. M., 1990. Receiving efficiency of monostatic pulsed coherent lidars. 1: theory. *Applied Optics*, **29**, 4111, doi:10.1364/AO.29.004111.
- Zhao, Y., Brewer, W. A., Eberhard, W. L., and Alvarez, R. J., 2002. Lidar measurement of ammonia concentrations and fluxes in a plume from a point source. *Journal of Atmospheric and Oceanic Technology*, **19**, 1928, doi:10.1175/1520-0426(2002)019<1928:LMOACA>2.0.CO;2.
- Zuber, M. T., Smith, D. E., Solomon, S. C., Phillips, R. J., Peale, S. J., Head, J. W., Hauck, S. A., McNutt, R. L., Oberst, J., Neumann, G. A., Lemoine, F. G., Sun, X., Barnouin-Jha, O., and Harmon, J. K., 2008. Laser altimeter observations from MESSENGER's first Mercury flyby. *Science*, **321**, 77, doi:10.1126/science.1159086.

Cross-references

Cryosphere, Measurements and Applications
 Ocean, Measurements and Applications
 Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties
 Optical/Infrared, Radiative Transfer
 Optical/Infrared, Scattering by Aerosols and Hydrometeors
 Radiation, Multiple Scattering
 Radiation, Volume Scattering

EMERGING TECHNOLOGIES, RADAR

Alina Moussessian
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Definition

To respond to the wide range of goals in Earth and Planetary sciences, radar systems must operate with a wide range of functionality. Parameters include multipolarization transmit and receive signals, multiband operation, adaptable resolution, scanning for increased swath, and very high measurement accuracy. In addition onboard data processing

techniques for data reduction and transmission and handling of large data sets are increasingly important as more complex, higher bandwidth radars are needed to meet mission requirements. Improved tools and support for warehousing and data mining are required to document the large number of missions and retain long duration data records required to perform key analysis, modeling, and monitoring.

Lowering mass and volume of radar systems is especially important for spaceborne missions both to lower spacecraft and launch vehicle costs and to enable combining multiple instruments such as radar/lidar on the same platform. Continued technology advancements focusing on risk and cost reduction are also critical.

Below are some of the technologies for future radar systems:

RF technologies

High-efficiency RF semiconductor devices are critical for future high-power systems. The use of wide bandgap semiconductors technologies such as GaN technology enables high-power, high-efficiency power amplifiers. For the most part, this technology can replace other devices, such as Si and GaAs, in power amplifier components.

In addition to semiconductor technology, some applications will still require the use of extended interaction klystron (EIK) and/or traveling-wave tubes (TWT). Further technology advancement in phase stable, higher peak power, higher efficiency, and longer life devices is needed.

Device technologies such as SiGe allow integration of RF and digital components onto the same chip enabling the concept of "radar on a chip." This is most beneficial for phased array applications where we need hundreds or thousands of these transmit/receive (TR) modules for electronic beam scanning.

In addition to high-efficiency, high-power RF transmitters, phase stable electronics both for receive and transmit are critical for future interferometry applications.

Quasi-optical techniques at millimeter-wave frequencies must be improved to support the front-end optics/electronics for millimeter-wave radars measuring crucial environmental parameters, such as clouds, aerosols, particulates, and other pollution.

Antenna technologies

In most cases the largest part of a radar is its antenna. Developing new materials and technologies for reducing antenna mass density both for passive and active antennas is critical. For reflector-type antennas, mesh technology is currently used at lower frequencies. Future radars will need larger mesh antennas at high frequencies (ka-band above). For these radars the development of multi-frequency shared feeds will allow sharing of resources for multiple measurements.

For very large active-phased arrays (size of 900 m² or greater), membrane technology is one way of reducing

mass density. Technology challenges in this area include integrating electronics with the array, manufacturing of these arrays, reliability, and the need for calibration to account for the antenna shape change.

MEMS technologies

MEMS resonators have the potential to incorporate low-loss front-end analog filtering into active antennas. This technology is becoming more important as the RF spectrum becomes more crowded, and digital processing is moving up the RF chain. As MEMS technology matures, other MEMS technologies such as MEMS switches might be usable in radar front-end electronics.

Digital technologies

Analog-to-digital converters (A/D) transform continuous analog radar signals into discrete digital signals. Extensive use of digital signal processing of radar data demands higher sampling rate A/D converters with larger dynamic range. Direct sampling of higher frequency signals, such as L and S band, requires faster A/D converters as well as higher dynamic ranges.

Due to further advancements in digital electronics, radars are using digital techniques in areas that RF technologies were previously used. One of these areas is the use of digital filtering, for example, for RFI mitigation. This allows the filtering of unwanted signals as the radar moves in space encountering different in-band interference. Another area where digital technology is replacing RF is digitizing higher frequency signals without having to downconvert the signal. For example, with current technology an L band (1.2 GHz) direct digital receiver is possible.

Another area which will dominate future radar applications is adaptive onboard science data processing. This along with development of a direct downlink capability provides near real-time data products tailored to users' needs.

Information system technology

We will need information technology advances for collecting, handling, and managing very large amounts of data and information in space and on the ground. These advanced information systems need to support new observations and information products; increase the accessibility and utility of science data; and reduce the risk, cost, size, and development time for future missions. These technologies are not unique to radar systems but apply to all sensors. However, radar systems typically collect large data which can take advantage of these processes. Technologies in this area include incorporating autonomy and intelligence within the sensing process and allowing rapid response to needed measurements, data acquisition on demand (in response to both science and society), sensor-to-sensor coordination, and interoperability. These technologies will increase the effectiveness of observing instruments or missions.

We also need technologies for combining observations of multiple Earth science variables into numerical models, data mining for information characteristics or content, dynamically acquiring and combining data from multiple data sources and distributed processing.

Constellations

Improved orbital control and tracking, along with satellite-to-satellite communication, is required to continue trends.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Kuhn, W., Mojarradi, M., and Moussessian, A., 2005. A resonant Switch for LNA protection in Watt-level CMOS Transceivers. *IEEE Transactions on Microwave Theory and Techniques*, **53**, 2819–2825.
- Moussessian, A., Del Castillo, L., Huang, J., Sadowy, G., Hoffman, J., Smith, P., Hatake, T., Derksen, C., Lopez, B., and Caro, E., 2005. An active membrane phased array radar. In *IEEE MTT-S International Microwave Symposium*, Long Beach, CA.
- Moussessian, A., Mojarradi, M., Johnson, T., Davis, J., Grigorian, E., Hoffman, J. P., Kuhn, W., and Caro, E., 2004. A completely integrated L- and S-band radar on a chip. Provisional Patent Application, NASA Case No. NPO 40869.
- Reuss, R. H., Chalamala, B. R., Moussessian, A., et al., 2005. Macroelectronics: perspectives on technology and applications. *Proceedings of IEEE*, **93**, 1239–1256.

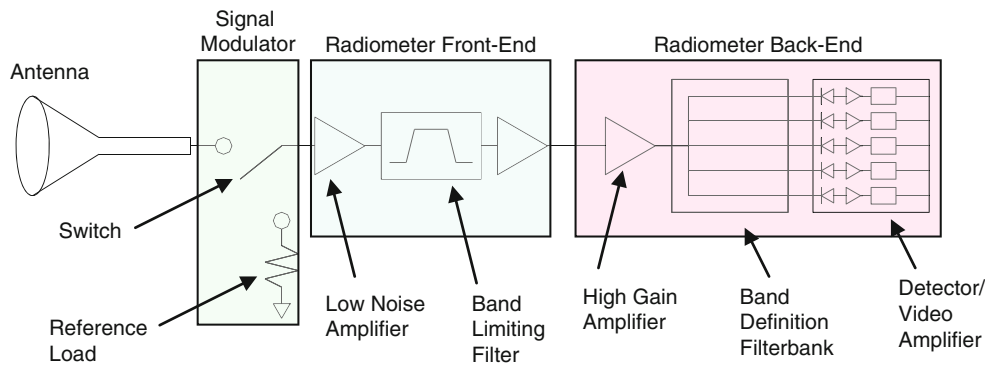
EMERGING TECHNOLOGIES, RADIOMETER

Todd Gaier

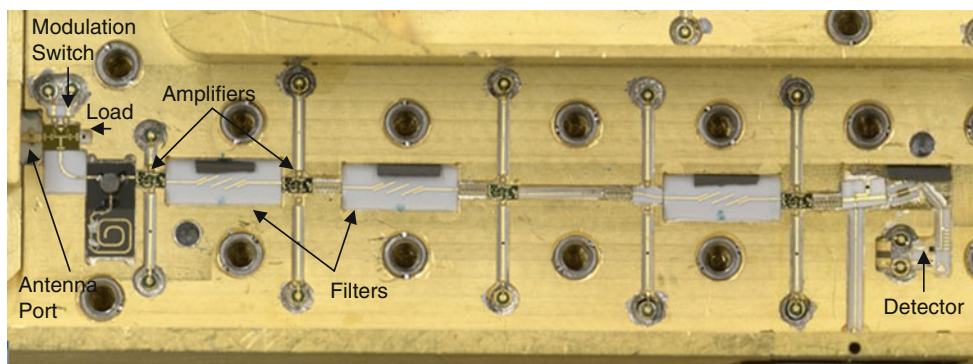
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition and introduction

Radiometric remote sensing involves the detection of emitted radiation and or reflected ambient radiation from an object. Such radiation typically originates as blackbody radiation from a remotely sensed object, undergoing partial reflection or absorption by intervening media. Detection is generally categorized into two types, *coherent detection* or *incoherent detection* (often referred to as direct detection). Coherent detection preserves the phase of the incoming wave, while incoherent detection does not. Examples of incoherent detectors include bolometers, thermopiles, CCD and APS sensors, as well as photoconductors and power detection diodes. Most passive remote sensing applications do not require phase preservation, although the advantages of phase-preserving detection make these systems the preferred choice for environmental applications in the frequency range



Emerging Technologies, Radiometer, Figure 1 Schematic of a basic radiometer suitable for remote sensing.



Emerging Technologies, Radiometer, Figure 2 Photograph of the Jason-II advanced microwave radiometer, 34 GHz channel, which shows all of the major components of a remote sensing radiometer (Courtesy D. Dawson).

1 GHz–1 THz. It is important to note that direct detectors do not have to impart additional noise to the incoming radiation, although practically speaking, cryogenic cooling of direct detectors is required to eliminate this added noise. When the observed object is very cold and the detected signals are very small (as in applications in radio astronomy), direct detection provides the lowest possible noise. The cooling requirement and the fact that most objects observed in remote sensing applications are already warm, resulting in a large inherent noise term, make these detectors less desirable than coherent detectors. For the rest of this entry, the discussion will be limited to coherent detection.

The basic coherent detection radiometer system can be divided into a front end and a back end. The front end amplifies the incoming signal (a voltage waveform) to a level where additional signal processing can occur without consideration of additional noise. The front end usually performs this function with the lowest possible additive noise available, but coherent systems always add some noise. The back end processes the amplified waveform extracting the desired information. Since most electronic components used in these systems exhibit inherent long-term instability, some sort of regular

calibration or signal modulation must be applied to the radiometer. A schematic of a basic radiometer is shown in Figure 1. A photograph of an integrated radiometer is shown in Figure 2.

Radiometer front ends

The choice of front-end technologies depends upon the application and the frequency of operation. Below 100 GHz low-noise amplifiers are readily available. Most modern amplifiers use solid-state semiconductor, three-terminal devices (transistors) as the active low-noise element. The development of these transistors has a remarkable history which is driven largely by demands in the computer and telecommunications industries. Common transistors can be classified into two types: bipolar transistors and field-effect transistors (FETs). One can think of a bipolar transistor as a current-controlled current source (i.e., the input current between one terminal and the second controls the current between the second and third terminal). The action of the current control can be used to make a small signal larger, referred to as *signal gain*. A FET is a voltage-controlled current source (i.e., the *voltage* between the first and second

terminals determines the *current* between the second and third). At frequencies below a few GHz, bipolar transistors offer excellent noise and are extremely stable over time. These devices are typically made of doped silicon or sometimes (gallium arsenide) GaAs. Recently, the computer industry has pushed the practical frequency of operation of bipolar transistors well beyond 10 GHz. This is largely due to the development of SiGe transistors. SiGe transistors provide competitive gain and noise up to 10 GHz (Weinreb et al., 2007), at costs of pennies per transistor. This technology is also important for receiver back ends as will be discussed later.

The other primary front-end amplifier technology is a type of FET called a high-electron-mobility transistor (HEMT). HEMTs are devices in which the semiconductor crystal structure is grown one atomic layer at a time. This confines the current carriers (either electrons or “holes”) to a two-dimensional surface, relatively free of scattering, providing a combination of high speed (with resulting high-frequency gain) and low noise (Duh et al., 1988).

A second key development in this field is the ability to create very small *gate* structures. The gate is the terminal of the FET which controls the current. In order to obtain very high switching speeds, the time for a carrier to transit the gate must be short. Electron-beam lithography allowed the creation of gate structures well under 1 μ in size (0.1 μ is readily achievable today). HEMTs using GaAs semiconductors have been used to build low-noise amplifiers for decades at frequencies up to 100 GHz. In the 1980s, HEMT technologies were employed in monolithic microwave integrated circuits (MMICs) in which several transistors were integrated onto single chips functioning as amplifiers. During the 1990s, experiments were performed with different materials systems. Indium phosphide (InP) emerged as a materials system with very high gain, low noise, as well as low operating power. InP transistors and MMICs are now regularly used in systems up to 100 GHz with *noise temperatures* below 300 K (Pospieszalski, 2005; Lai et al., 2005).

Recent developments in this field have shown that it is still improving. Techniques have become available to reduce gate sizes as small as 35 nm. InP MMIC amplifiers have now been demonstrated at frequencies as high as 350 GHz (Pukala et al., 2008). These very high operational frequencies had been impossible for devices with gain. For remote sensing applications above 100 GHz, it has been typically required to utilize *heterodyne detection*. In a heterodyne receiver, the incoming signal is fed into a nonlinear device (usually a diode) where it is *mixed* with a sinusoidal voltage at almost the same frequency (local oscillator or LO frequency). A waveform results which has an amplitude proportional to the incoming signal but at a frequency which is the difference between the signal and LO frequencies. The heterodyne architecture is extremely useful if the observed object has high-spectral-resolution features, such as high-altitude atmospheric molecular line emission and absorption (Waters et al., 2006).

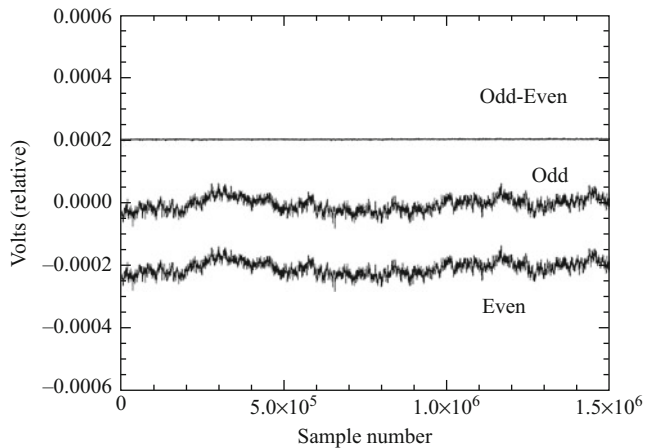
Diode mixers are the dominant radiometer front end between 100 GHz and 1 THz. Most of these diodes are Schottky diodes, formed when a metal and a semiconductor are brought into contact. Schottky diode mixers have been built up to several THz and have been used extensively in astronomy and atmospheric chemistry. The earliest mixers were built by manually pressing a metal whisker onto a piece of doped semiconductor. While these devices were used at very high frequencies, there were obvious reliability issues. Semiconductor and lithographic processes, similar to those used for transistors, are now used to make planar diode structures, also operating at THz frequencies. These planar devices can even be bought commercially. To obtain the lowest possible noise, mixers can be fabricated from superconducting devices called superconductor-insulator-superconductor or SIS mixers. While these devices have noise limited only by quantum mechanics, they require cryogenic cooling, making them impractical for many remote sensing applications. SIS mixers are widely used in radio astronomy (Pan et al., 1989).

Local oscillator technologies are a key part of heterodyne receiver front ends. Detection of signals at several hundred GHz requires oscillators working at these frequencies. For a long time, the primary device used was the Gunn-effect diode. Gunn diodes have a negative resistance, that is, as the voltage is decreased, the current increases. Such a device is inherently unstable. When coupled to a resonant cavity, this instability leads to oscillation, with the frequency of oscillation dependent upon the size of the cavity. Gunn oscillators capable of generating tens of megawatts have been built up to 160 GHz. To obtain even higher frequencies, diode frequency multipliers are employed.

Advances in power amplifier technology have enabled power generation at ever-increasing frequencies. GaAs power amplifiers are capable of generating hundreds of mW at frequencies above 100 GHz. These amplifiers have largely displaced Gunn oscillators as the primary source of LO power in mm-wave radiometers.

Calibration and modulation

The inherent gain instability of high-frequency circuits creates the need to stabilize the radiometer. The most common form of signal stabilization is the Dicke-switched radiometer (Dicke, 1946). The Dicke radiometer has as a first element, a modulation switch which rapidly exchanges the input signal with a stable reference load. The output of the radiometer is then a synchronous signal representing the remote sensed object and the reference load. Differencing of these signals on short timescales provides a calibration on a timescale shorter than the receiver gain can vary. A graphical depiction of this is shown in Figure 3. Many variations of modulated radiometers exist, including noise-adding radiometers (in which a noise source is added synchronously to provide gain calibration) and correlation radiometers (which allow for simultaneous



Emerging Technologies, Radiometer,

Figure 3 Pseudocorrelation Radiometer output. A time series (roughly 1 h) of data are shown with the radiometer alternately viewing a “remote sensed target” (*even*) and an internal reference load (*odd*). Each of the data sets displays long-term noise (due to gain instability). The differenced data are also shown, with considerably lower resulting noise and greatly improved stability. This demonstrates the power of signal modulation in a radiometer in removing time-correlated excess noise (Graphic courtesy M. Seiffert).

detection of two input signals through a common signal path). For each of these systems, signal modulators are required.

In order to be effective, switches used for Dicke modulation are usually in front of the first gain element. Historically, switches have been mechanical, ferrite, or semiconductor diode switches. Because of the high speeds and reliability required, mechanical switching is now limited to radiometers which can have the entire optical and receive system quickly scanned across the observed field of view. Ferrite switches use the time asymmetric properties of magnetic fields to switch signals between two ports of a three-port switch. The magnetic field of a circulating switch is typically provided by an electromagnet. Reversing the current in the electromagnet reverses the magnetic field which switches the microwave signal at a common port. In recent years, diode switches have been used for signal modulation. These switches take advantage of the conducting versus nonconducting states of a diode depending upon their *bias* or DC current. PIN (*p*-intrinsic-*n* doped semiconductor) diodes are particularly useful as switches as they are small, low loss (at low frequencies), support fast switch rates, and are repeatable and reliable with no moving parts.

PIN switches can also be used to make phase modulators. Phase modulators are useful in providing radiometer stability in a correlation (or pseudocorrelation) radiometer. In a correlation radiometer, couplers are used to provide a common signal path for incoming signals and the stable

reference. While they do share a signal path, these signals must be separated prior to detection, which creates residual instability. Phase modulation stabilizes this residual by switching the remote sensed and reference signals prior to detection. The advantage of such a system is that the modulation element need not be located prior to the first gain stage, which in turn means that signal loss is not a primary concern. PIN diodes tend to have high loss at high frequencies. For this reason, high-frequency radiometers often use correlation techniques for stabilization (Jarosik et al., 2003).

The noise-adding radiometer is also used to stabilize a signal. While it seems counterintuitive to add noise to a receive system, it is an effective means of providing necessary gain stabilization. In this system, a noise source is coupled into the receiver input (with low efficiency) and is rapidly modulated. The noise source is typically a noise diode (similar to a Zener diode), which is operated alternately in the off state and in a state near avalanche. In avalanche, the diode carriers undergo ballistic collisions with the lattice to produce excess noise. This white noise looks like a blackbody of high temperature (typically thousands of degrees kelvin). By coupling in a small portion of the signal, a stable offset of tens of kelvin can be generated, which, while small compared to the receiver noise, is large enough to provide long-term gain stabilization. Because of the statistics of the noisy signal, gain adding alone is typically insufficient to provide stability and is typically used with some other form of signal modulation.

Radiometer back ends

The back end receives the amplified input signals and performs necessary processing. In the simplest form, the back end is a band definition filter and a power detector. Such a system measures the *radiometric temperature* in a band. Power detectors for remote sensing systems are almost always either Schottky barrier diodes or tunnel diodes. The diode rectifies the incoming signal with a DC voltage response that, to first order, is proportional to the square of the voltage or the power. In order to rectify such high-frequency signals, it is necessary for the diode to have a fast response. Because Schottky diodes can be fabricated with small features, they are typically used for detection applications at frequencies above 10 GHz. Limitations of the diodes include post-detection instability and significant temperature dependence. For this reason, tunnel diodes are the preferred choice below 10 GHz. There have been recent developments in high-frequency tunnel diodes (Moyer et al., 2008), although little data exist for their application in remote sensing. Detector diodes typically output signals at the 1 μ V level in order to maintain linearity. The noise fluctuations about this signal are typically three to five *orders of magnitude smaller* than this, requiring low-noise *video* amplification (the term *video* is typically used to denote the post-detection signal).

Modern back-end processors take advantage of emerging digital technologies. Some of these back-end processors include *spectrometers*. A spectrometer takes the amplified input signal (or mixed signal) and processes the electromagnetic spectrum to determine the power in discrete frequency bands. As digital signal processing speeds improve, more technologies are becoming available for remote sensing applications. Digitizers with speeds faster than 2 GS/s are now commercially available. Processors also exist to handle these data rates. Field-programmable gate arrays (FPGAs) have been used to process these high-speed signals both in time and frequency domain. Spectrometers have been demonstrated using Fourier transform and polyphase filter bank technologies (Parsons et al., 2006).

The same technology can also be used to purge data of radiofrequency interference (RFI). RFI is becoming a greater problem for remote sensing applications in the frequency range 1–15 GHz as commercial communications technologies become widespread. Even with protected bands, contamination of the electromagnetic spectrum requires active monitoring and vetoing by current remote sensing experiments. Fortunately, the statistics of RFI signals differ from thermal signals, allowing for high-speed processing to flag and remove contaminated data (Ruf et al., 2006).

The final type of back-end digital processor is the cross correlator. The correlators multiply signals from one receiver with another. These signal products can be used in aperture synthesis interferometry. Aperture synthesis radiometry has become more widely used in recent years as large apertures without moving parts are becoming required to meet remote sensing requirements for soil moisture and ocean salinity measurements as well as continuous observations of atmospheric (SMOS, GeoSTAR).

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Bibliography

- Dicke, R. H., 1946. The measurement of thermal radiation at microwave frequencies. *The Review of Scientific Instruments*, **17**, 268–275.
- Duh, K. H. G., Pospieszalski, M. W., Kopp, W. F., Ho, P., Jabra, A. A., Chao, P.-C., Smith, P. M., Lester, L. F., Ballingall, J. M., and Weinreb, S., 1988. Ultra-low-noise cryogenic high-electron-mobility transistors. *IEEE Transactions on Electron Devices*, **35**(3), 249–256.
- Jarosik, N., et al., 2003. Design, implementation and testing of the MAP radiometers. *The Astrophysical Journal Supplement Series*, **145**, 413.
- Lai, R., Grundbacher, R., Sawdai, D., Uyeda, J., Biedenbender, M., Barsky, M., Gutierrez-Aitken, A., Cavus, A., Chin, P., Liu, P. H., Bhorania, R., Streit, D., and Oki, A., 2005. Production InP MMICs for low cost, high performance applications. In *International Conference on Indium Phosphide and Related Materials*, May 8–12, 2005, pp. 598–602.

- Moyer, H. P., Schulman, J. N., Lynch, J. J., Schaffner, J. H., Sokolich, M., Royter, Y., Bowen, R. L., McGuire, C. F., Hu, M., and Schmitz, A., 2008. W-band Sb-diode detector MMICs for passive millimeter wave imaging. *IEEE Microwave and Wireless Components Letters*, **18**(10), 686–688.
- Pan, S.-K., Kerr, A. R., Feldman, M. J., Kleinsasser, A. W., Stasiak, J. W., Sandstrom, R. L., and Gallagher, W. J., 1989. An 85–116 GHz SIS receiver using inductively shunted edge junctions. *IEEE Transactions on Microwave Theory and Techniques*, **37**(3), 580–592.
- Parsons, A., Backer, D., Werthimer, D., and Wright, M., 2006. A new approach to radio astronomy signal processing: packet switched, FPGA-based, upgradeable, modular hardware and reusable, platform-independent signal processing libraries. In *Proceedings USNC-URSI*, Boulder, CO, January 2006.
- Pospieszalski, M. W., 2005. Extremely low-noise amplification with cryogenic FETs and HFETs: 1970–2004. *IEEE Microwave Magazine*, **6**(3), 62–75.
- Pukala, D., Samoska, L., Gaier, T., Fung, A., Mei, X. B., Yoshida, W., Lee, J., Uyeda, J., Liu, P. H., Deal, W. R., Radisic, V., and Lai, R., 2008. Submillimeter-wave InP MMIC amplifiers from 300–345 GHz. *IEEE Microwave and Wireless Components Letters*, **18**(1), 61–63.
- Ruf, C. S., Gross, S. M., and Misra, S., 2006. RFI detection and mitigation for microwave radiometry with an agile digital detector. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(3), 694–706.
- Waters, J. W., et al., 2006. The earth observing system microwave limb sounder (EOS MLS) on the aura satellite. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(5), 1075–1092.
- Weinreb, S., Bardin, J. C., and Mani, H., 2007. Design of cryogenic SiGe low-noise amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, **55**(11), 2306–2312.

EMERGING TECHNOLOGIES, SENSOR WEB

Mahta Moghaddam¹, Agnelo Silva¹ and Mingyan Liu²
¹Electrical Engineering – Electrophysics, University of Southern California, Los Angeles, CA, USA
²Electrical and Computer Engineering, University of Michigan, Ann Arbor, MI, USA

Synonyms

Environmental sensor network; ESN; Wireless sensor network; WSN

Definition

Sensor webs are networks of sensors that can operate in a coordinated fashion, for example, through centralized or decentralized control centers. The controller can command the sensor nodes to modify their measurement schedule or configuration in response to environmental factors or to achieve certain measurement goals. Sensor webs are envisioned as a means of providing spatially and temporally adaptive in situ measurements for validation of remote sensing observations.

Introduction

Satellite remote sensing often results in data and retrieved geophysical products with resolutions that are significantly coarser than the scale of variations of the phenomena they represent. As an example of a geophysical variable, soil moisture fields retrieved from satellite observations using microwave instruments have resolutions on the order of kilometers if not coarser. However, the soil moisture fields themselves have spatial dynamics at the scales of several meters (e.g., because of topographic and land-cover variations), hundreds of meters (e.g., because of land-cover and soil texture variations), and kilometers (e.g., because of precipitation). Therefore, the coarse-resolution retrievals at the satellite pixel scale may not accurately represent the true mean of the soil moisture field.

The validation of satellite retrievals is therefore a challenging task. It requires the use of in situ sensor networks, whose node placement has to be such that the proper spatial statistics are represented and such that the in situ estimate of the mean soil moisture within the coarse-resolution pixel is close to the true mean. Furthermore, the measurement schedule of the sensor nodes has to be such that the temporal variations of soil moisture are properly captured. At the same time, it has to be dynamic so that, if necessary, only the minimum number of measurements is taken so that energy usage is optimized by the network.

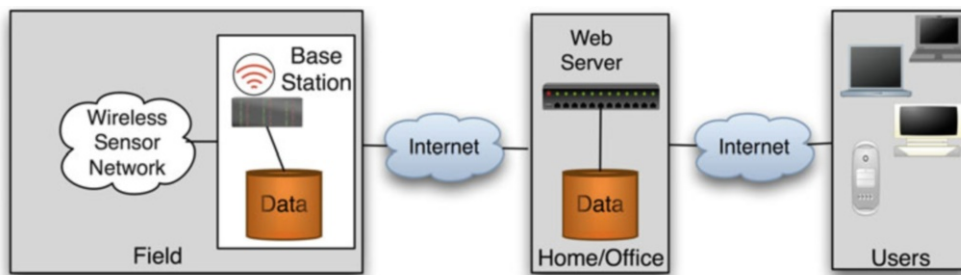
A large number of sensor nodes may be necessary to sufficiently represent the spatial variations of the coarse-resolution field. For example, for the NASA Soil Moisture Active and Passive (SMAP) satellite mission, scheduled for launch in 2014, the primary soil moisture product will have a resolution of 10 km (Entekhabi et al., 2010). Depending on the topographic, vegetation, and soil texture heterogeneity present in various pixels, tens or hundreds of sensors might be needed for proper retrieval validation. However, it is not feasible to deploy such large number of sensor nodes with conventional data loggers and manual data collection schemes. Instead, wireless network concepts have to be used that allow near real-time data upload. There are many challenges with such large-scale outdoor networks, including energy management, lifetime, environmental robustness, network capacity, and costs. Such challenges

are illustrated by the following discussion involving the design of the sensor web for the SMAP mission.

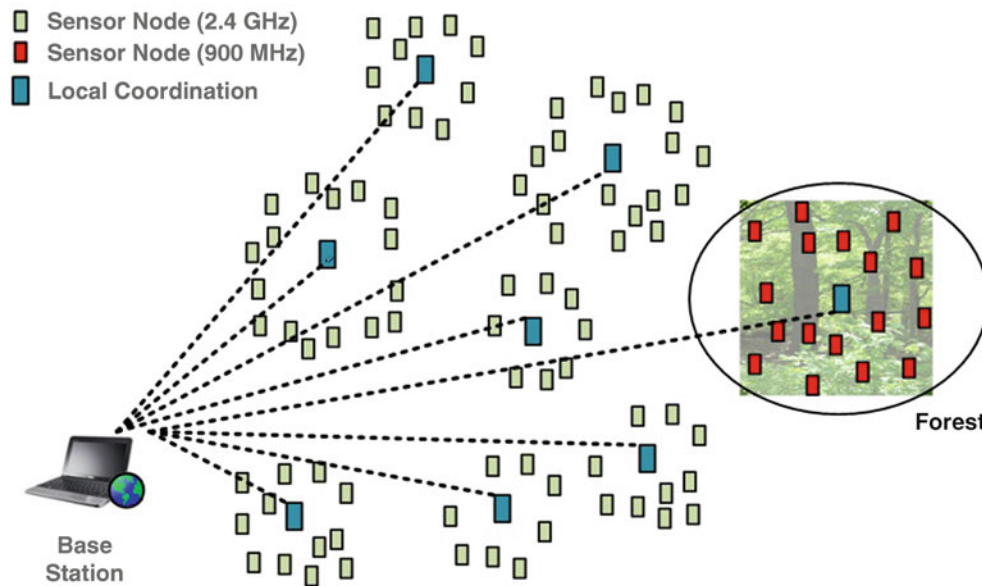
Global architecture of the sensor web

The global architecture of the in situ sensor network is shown in Figure 1. The system consists of a field element and a home/office element. A wireless sensor network is deployed over a target field, along with a base station that performs data collection and sensing control, and a database collocated with the base station for local data storage. At each sensing site, several sensors (for example, soil moisture sensors) are deployed at different depths and connected to the sensor node (i.e., a ground wireless module). The base station receives data wirelessly from each sensor node; it can also control their measurement schedules on demand. It also periodically (every half an hour) uploads the collected sensor data through a long-range link, such as a 3G connection, to a database server located in the home base.

Initially, this network was built using the ZigBee (IEEE 802.15.4 plus higher layer specifications) standard (Moghaddam et al., 2010). The first deployment of this network was in a field near Canton, Oklahoma. ZigBee was chosen as it allowed the formation of a multi-hop network and due to its mature technology that could significantly shorten the development and production cycle. The disadvantage was that a router node under the ZigBee specification could not be put to sleep mode, thus consuming significantly more energy and requiring large batteries and large solar panels. It also was proven to be rather unstable for outdoor field environment due to poor router-base station connection, causing end devices (or nodes) to switch parent-child association. Therefore, the architecture was modified to adopt a two-tiered hierarchy: the lower layer consists of a local coordinator (LC) node and multiple sensor nodes or end devices (ED) associated with the LC node. The upper layer consists of LC node(s) and a base station. In contrast to existing network routers, the LC node can sleep and its energy requirements are significantly reduced. The LC node may be equipped with two radio interfaces, allowing it to communicate within the two layers using different radio technologies, effectively making the two layers logically separated. The lower layer uses the IEEE 802.15.4 standard but not



Emerging Technologies, Sensor Web, Figure 1 Global architecture of the wireless sensor network system.



Emerging Technologies, Sensor Web, Figure 2 A large wireless sensor network may have multiple local coordinators servicing a variety of different landscape types. Choice of local transceivers should be transparent to the network for extensibility.

the ZigBee suite. The advantages of this design include (1) flexibility in developing an open protocol on top of 802.15.4 for the lower layer and multiple candidate solutions for the upper layer; (2) the logical separation between the two layers, making sleep scheduling of the ED nodes much easier to control; (3) a different radio solution for the upper layer, allowing the system to span over much longer distances; and (4) ease of scaling up of system architecture. Using this design, the network can be scaled up (Figure 2) to contain multiple local coordinators that can be deployed in multiple landscape types and to cover a span of several kilometers.

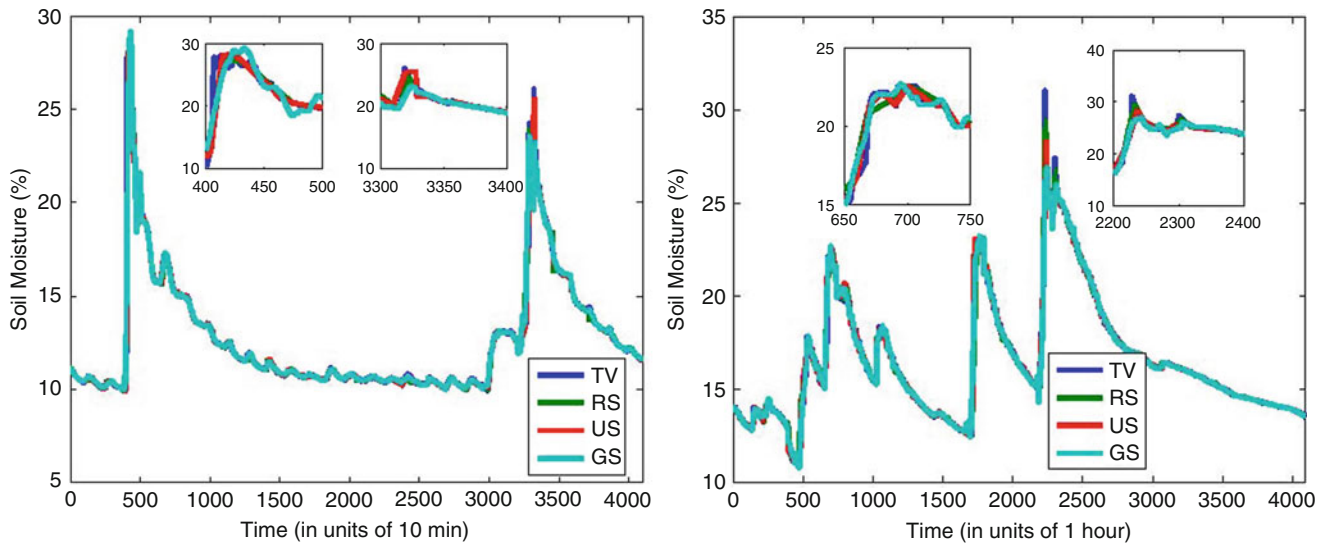
Sensor placement, scheduling, and field mean estimation

The true mean of geophysical variables such as soil moisture fields is a function of time and of the state of the soil surface. Its determination ideally would require a very fine sampling of the remote sensing satellite footprint, both spatially and temporally. This, however, is cost prohibitive; manually installing these sensors is expensive, and their battery power does not allow us to continuously sample, as we need them to last a reasonably long period of time (months or even years). These considerations pose severe limitations on how many sensors can be made available and how frequent they can be used/activated. The overall objective is thus to place and activate sensors such that the field mean may be estimated to a desired accuracy subject to budgetary constraints, e.g., the total number of sensors available, the total amount of available energy at each sensor, and bandwidth (Wu and Liu, 2012a).

There are two elements to the above problem: one is the determination of the best set of locations within the sensing field to place a limited number of sensors (sensor related cost constraint) and the other is the optimal dynamic operation of these sensors (when and which to activate) once they are placed (energy constraint).

These two elements are coupled. For instance, if energy of operation is a bigger concern than placement costs, then one can choose to place more sensors to compensate for a desired, reduced sampling rate. The reverse holds as well. In addition, activation and sampling decisions can influence where sensors should be placed and vice versa. But jointly considering and optimizing both elements leads to a problem whose complexity is prohibitive both analytically and computationally.

For a given placement plan, the sensor measurement scheduling problem becomes one that aims to minimize the estimation error (for the reconstructed soil moisture process using measured samples) subject to a certain energy/sampling rate constraint, or to minimize the sampling rate subject to an estimation accuracy criterion, or to minimize certain weighted sum of both. The estimation can be done both in a closed-loop fashion and an open-loop fashion. Under a closed-loop approach, temporal and spatial statistics of the soil moisture evolution is first learned using training data (either real or simulated). This knowledge, which describes what is likely to happen given what has already happened, together with samples already collected, i.e., what we know has happened, can help predict the future and thus make judicious decisions on the best time to take the next measurements. This is essentially the idea behind a partially observed Markov decision process (POMDP) formulation of this problem (Shuman et al., 2010).



Emerging Technologies, Sensor Web, Figure 3 Recovery accuracy of compressed sensing approach under three types of measurement schedules (US: uniform/periodic sampling; RS: random sampling; GS: Gaussian sampling; TV: true value).

Under an open-loop approach, recent results from the theory of compressive sensing can be applied. This technique exploits an underlying sparsity feature of the measured signal and is able to reconstruct the soil moisture process from a very small number of samples to great accuracy. Advantages of this approach include the following: (1) it does not require training or a priori statistical knowledge of the soil moisture process, (2) the sampling sequence (measurement times) can be completely determined offline and in its simplest form can be a periodic sequence, therefore making implementation very easy, and (3) the same signal reconstruction technique can be used if we augment the sampling sequence with exogenous information like rainfall (e.g., take more samples during a rainfall event and less during dry periods). The figures below show the recovery accuracy of this approach under three types of measurement schedules (US, uniform/periodic sampling; RS, random sampling; GS, Gaussian sampling; TV, true value) (Wu and Liu, 2012b) (Figure 3).

Energy management schemes

Flexible energy management solutions are needed to meet the lifetime requirements of remote sensing missions such as SMAP. In doing so, increased reliability and reduced cost of wireless sensor network operation, extensibility to a large variety of target environments, and therefore an open and generalized architecture solution are needed.

Past experience has shown a significant number of battery failures caused by extreme temperatures as shown in Figure 4. Therefore, for many realistic deployment



Emerging Technologies, Sensor Web, Figure 4 Sensor nodes based on solar panels and rechargeable batteries typically have problems associated with sub-zero temperatures.



Emerging Technologies, Sensor Web, Figure 5 A Ripple-2 node powered by non-rechargeable batteries (top) installed in SoilSCAPE field locations such as Canton, OK, USA (bottom).

scenarios, the estimated lifetime of two or more years for a sensor node cannot be realized with rechargeable batteries because of their temperature sensitivity. Non-rechargeable batteries are the best option in this regard because of their tolerance to both high and low temperatures. If the usage of the network is consistently maintained low, such an approach could be very cost-effective and robust. For instance, assuming measurements every 15 min, the batteries need only be replaced every 2–4 years (depending on the model/type of the battery). No external part, such as a solar panel, is required to be exposed to the environment. This solution has been implemented in the “Ripple-2” system, developed by the authors (Figure 5) as part of the Soil moisture Sensing Controller And oPtimal Estimator (SoilSCAPE) project.

However, it is not always possible to maintain a low duty cycle. For instance, one could expect to periodically

(and temporarily) modify the schedule of the sensor nodes to make measurements with the full node-density capacity of the network to optimize the scheduling and estimation processes. The main drawback in this case is the high energy cost. By adopting an energy-harvesting model, such as solar panels, one could compensate the additional loss of energy. Accordingly, two additional versions of Ripple-2 node are developed. The Ripple-2B is based on supercapacitors and a solar panel, and the Ripple-2C is a hybrid solution with supercapacitors and non-rechargeable batteries.

The usage of non-rechargeable batteries is not the preferred option for conventional sensor web solutions, in particular for outdoors. The main reason is the high cost involved in frequent battery exchanges (Weddell et al., 2008). Also, typically the energy consumption of the nodes is not uniform among them, and multiple

maintenance trips are necessary. Under the Ripple-2 development, the reasons behind this scenario are investigated, and it is concluded that cooperation among nodes is the main factor for the quick battery depletion. In other words, the average energy spent by a node *serving* the network node is multiple times the energy spent by a node to periodically take measurements and transmit data to the base station (Silva et al., 2012a).

The authors envisioned a scenario where the network is segmented into physical clusters and each segment has a data collector node (LC) that can communicate with all the sensor nodes in that segment by means of a single hop (direct link) (Silva et al., 2012b). By carefully assigning the location of the LC in relation to the sensor nodes of that area, the collaboration among sensor nodes due to multi-hopping is not necessary. As a result, the network overhead is drastically reduced from typical 3–10 % to less than 1 %. Also, a sensor node does not need to periodically wake up to *serve* the network (i.e., relaying messages). Therefore, the sleep scheduling of a Ripple-2 nodes is only governed by the application duty cycle. In other words, the energy consumption of a node has a very deterministic behavior, and energy balance among nodes is finally achieved no matter the size of the network.

Because a Ripple-2 node can potentially have a very long and continuous sleeping period (i.e., hibernation mode), it is possible to turn off many modules, such as the radio transceiver, rather than just put these modules into standby mode. The energy savings achieved by following this approach can be as high as 50 % (Silva et al., 2012a). Combined with the savings related to a reduced network overhead, the lifetime of a non-rechargeable battery of a Ripple-2 node can be realistically extended by multiple folds assuming a low duty-cycle regime, such as soil moisture measurements every 20 min (Menachem and Yamin, 2004).

However, there is an important trade-off in relation to non-rechargeable batteries: the *pulse current effect* which is the nonlinear and drastic energy capacity reduction of a battery when it is discharged by a slight higher current, such as 50 mA (Silva et al., 2012b). No matter how quick is this pulse current (e.g., a very fast radio transmission), the energy capacity of the battery is impacted and its lifetime can be as small as 10 % of the nominal/expected lifetime (Silva et al., 2012a). In order to solve this issue, Ripple-2 adopts supercapacitors as *power-matching* components. That is, these capacitors are slowly charged by a low current from the battery and quickly discharged by a high current (radio). Such approach protects the battery against the pulse current effect but also creates additional data latency issues when the node needs to transmit data very frequently. However, it is not a problem for a low duty-cycle web sensors used in SMAP.

Summary

To respond to the challenge of validation of coarse-scale remote sensing retrieval products such as soil moisture,

a generalized wireless sensor network architecture has been developed. This architecture can be scaled to large-scale outdoor wireless sensor webs with flexible placement, scheduling, and power management schemes. The latest implementation of this architecture has been termed “Ripple-2” (with Ripple-1 being the first generation of this architecture). Due to its advantages, this architecture can be extended not just for soil moisture but for other sensing applications by making it flexible enough for other processor platforms and wireless technologies. Such applications can include any environmental monitoring application that has an extensive network deployed over a large area.

The high degree of robustness, energy efficiency, and reliability of Ripple-2 are achieved under the assumption of low duty cycles (e.g., sensor measurements every 10–20 min) and data latencies from seconds to minutes.

The Ripple-2 architecture can be considered a milestone in wireless sensor networks (WSNs) because of its specialization for low duty cycle and data-centric applications, breaking well-established concepts for WSNs. Without increasing the costs, the energy performance of Ripple-2 nodes is significantly superior compared to any similar WSN/telemetry solution. In fact, even non-rechargeable batteries can now be considered as a cost-effective option. However, technological enhancements can provide the path to turn Ripple-2 into a *generic* WSN solution.

Bibliography

- Entekhabi, D., Njoku, E., O'Neill, P., Crow, W., Entin, J., Jackson, T., Johnson, J., Kimball, J., Koster, R., McDonald, K., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., and Tsang, L., 2010. The soil moisture active and passive (SMAP) mission. *Proceedings of the IEEE*, **98**(5), 704–716 (featured on cover).
- Menachem, C., and Yamin, H., 2004. High-energy, high-power pulses plus battery for long-term applications. *Journal of Power Sources*, **136**, 268–275.
- Moghaddam, M., Entekhabi, D., Goykhman, Y., Li, K., Liu, M., Mahajan, A., Nayyar, A., Shuman, D., and Teneketzis, D., 2010. A wireless soil moisture smart sensor web using physics-based optimal control: concept and initial demonstration. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **3**(4), 522–535.
- Shuman, D., Nayyar, A., Mahajan, A., Goykhman, Y., Li, K., Liu, M., Teneketzis, D., Moghaddam, M., and Entekhabi, D., 2010. Measurement scheduling for soil moisture sensing: from physical models to optimal control. *Proceedings of the IEEE*, **98**(11), 1918–1933.
- Silva, A. R., Liu, M., and Moghaddam, M., 2012a. Power Management Techniques for Wireless Sensor Networks and Similar Low-Power Communication Devices Based on Non-Rechargeable Batteries. *Journal of Computer Networks and Communications*, 2012 (article ID 757291), 10 pages, doi:10.1155/2012/757291.
- Silva, A., Liu, M., and Moghaddam, M., 2012b. Ripple-2: A non-collaborative, asynchronous, and open architecture for highly-scalable and low duty-cycle WSNs. In *Proceedings ACM International Workshop on Mission-Oriented WSN (MiSeNet' 12)*. Istanbul.

- Weddell, A. S., et al., 2008. Alternative energy sources for sensor nodes: rationalized design for long-term deployment. In *Proceedings IEEE International Instrumentation and Measurement Technology Conference (IMTC' 08)*. Victoria, British Columbia, pp. 1370–1375.
- Wu, X., and M. Liu, 2012a. In-Situ Soil Moisture Sensing: Optimal Sensor Placement and Field Estimation. *ACM Transactions on Sensor networks*, vol. 8, no. 4, pp. 33:1–33:30.
- Wu, X., and M. Liu, 2012b. In-situ soil moisture sensing: measurement scheduling and estimation using compressive sensing. *International Conference on Information Processing in Sensor Networks (IPSN)*. Beijing.

ENVIRONMENTAL TREATIES

Alexander de Sherbinin
Center for International Earth Science Information
Network (CIESIN), Columbia University,
Palisades, NY, USA

Synonyms

Multilateral Environmental Agreements (MEAs)

Definition

Environmental treaty: an environmental treaty is an agreement under international law entered into by two or more states addressing natural resource management or environmental protection. A treaty may also be known as an agreement, protocol, covenant, convention, exchange of letters, accord, exchange of notes, or a memorandum of understanding.

Introduction

Over the past 40 years, there has been a parallel increase in the number, range, and complexity of environmental agreements and the number and sophistication of remote sensing instruments. There were 172 environmental treaties in 1970, and that number has risen to more than 475 today (Mitchell, 2008), covering almost every conceivable environmental system. In terms of civilian or commercial remote sensing instruments, in 1970 there were only a handful of meteorological satellite sensors, whereas today there are more than 100 operational sensors with a wide range of spatial, spectral, and temporal resolutions.

Until the 1990s, the two communities – the international environmental law community, on the one hand, and remote sensing scientists, on the other – largely operated in isolation from one another. True, in the 1980s, images from the Total Ozone Mapping Spectrometer (TOMS) instrument onboard Nimbus-7 were used to document the seasonal depletion of ozone over the Antarctic, a visualization capability that significantly contributed to the Montreal Protocol of the Vienna Convention on Ozone Depleting Substances. But no one had asked the question: What kind of remote sensing data

products are required by the environmental treaty community? In the late 1990s, the two communities began to assess – through a series of workshops and dialogues – the ways in which remote sensing could contribute to international environmental policy making. This entry addresses a number of the opportunities and constraints for the application of remote sensing data to environmental treaties that have been identified through recent dialogues and documented by research projects and publications (cf. CIESIN, 2001; de Sherbinin, 2005; Johnston, 2006; Strand et al., 2007).

Opportunities

One of the greatest attributes of remote sensing is its synoptic view with wall-to-wall coverage. This, coupled with the fact that the data are “objective” and consistent over time, at least compared to the highly variable in situ monitoring between countries and over time, makes remote sensing a potentially ideal source of data for a range of environmental issues of international concern. These include habitat loss, biodiversity conservation, desertification, transboundary air pollution transport, eutrophication of coastal waters, and greenhouse gas emissions from land-based sources, among others. The data are also applicable for regions of widely varying scales, from global down to transboundary resource management contexts.

Another advantage is that, unlike most other data gathering techniques, collection of remote sensing imagery does not infringe national sovereignty. Furthermore, it enables data collection in remote, inaccessible, or war-torn locations at much lower costs than ground-based methods and without risk to personnel.

In terms of treaty processes, remote sensing can be applied in a number of areas. Perhaps the most common use is in the context of environmental assessments. Both the Intergovernmental Panel on Climate Change (IPCC) and the Millennium Ecosystem Assessment relied heavily on remote sensing data to reach science-based conclusions about the scope of climate change and the loss of ecosystem services, respectively. Information from scientific assessments often feed directly into treaty negotiations, leading to new protocols or amendments to treaty texts. Another area of great potential is in national reporting, an often data-intensive process that can be greatly facilitated by remote sensing. For example, in the second national reports of the Convention on Biological Diversity, contracting parties were asked if they were using rapid assessment or remote sensing techniques. Thirteen parties reported a lot of use, 58 reported some use, and 33 reported no use or that they were exploring the possibility.

Perhaps the most important role of remote sensing is its support for the broader political process. Images capture the significance of a problem in a way that the written word does not. Remote sensing before and after images of deforestation or deglaciation has served to bolster public support for stronger environmental agreements

and the national laws that contribute to treaty implementation. Remote sensing has illustrated in powerful ways alterations to a wide range of ecosystems such as the Everglades in the USA, the Aral Sea, and the Amazon basin (UNEP, 2005). These same images can also serve as wake-up calls to political leaders. Political support at the highest levels for the Mesoamerican Biological Corridor, an agreement under the Central American Commission for Environment and Development (CCAD), can be partly attributed to a Landsat image of the Mexico-Guatemala border that portrayed the extent of deforestation in the Mexican state of Chiapas as compared to the neighboring Guatemalan Petén (de Sherbinin et al., 2002).

Traditionally, one of the problems of remote sensing was its high cost. The costs of imagery, software packages, computer processing, and trained technicians were beyond the reach of many developing countries and represented significant barriers even for many developed countries. Fortunately, this is beginning to change. In the case of imagery, costs have come down considerably for all except the highest resolution (1–4 m) commercial imagery. MODIS data are free of charge, Landsat imagery will soon be available at no cost, and data from many other instruments, such as ASTER, are freely available for scientific uses. In addition, there has been a proliferation of instruments with widely ranging characteristics – including radar instruments – that cover a much wider spectrum of the Earth and environmental processes than the few land and meteorological satellites of the 1980s. The growth in availability of imagery with different characteristics has inevitably brought costs down, even for commercial imagery.

The cost of commercial software packages is still fairly high, though a number of shareware packages are now available with some impressive capabilities. The American Museum of Natural History's Biodiversity Informatics web site lists 12 free software packages (AMNH, undated). There are others, such as TerraLook, which prepackages ASTER imagery for use in assessing land cover change in and around protected areas. Although they lack analytical capabilities, web-based applications such as Google Earth and NASA World Wind are popularizing remote sensing imagery in ways that can only enhance their application in the environmental policy realm.

Computer processing power has increased exponentially, and costs have come down in that domain as well. Probably, the only area that has not seen significant cost declines is the salaries for trained technicians. We return to the issue of training in the next section.

In terms of institutional support for the application of remote sensing to environmental treaties, a great stride forward has been made with the establishment of the Group on Earth Observations' (GEO) Global Earth Observation System of Systems (GEOSS). GEOSS is facilitating exactly the kind of dialogue between data users in the environmental policy realm and remote sensing scientists that began in the 1990s, and the dialogue is

now being expanded to Internet technologists who are dreaming up new ways to combine remote sensing imagery with in situ data to better understand and manage the environment.

Constraints

As mentioned in the previous section, a number of former barriers to the use of remote sensing for international environmental policy – such as the high cost of imagery and software – have fallen. Nevertheless, there remain some technical hurdles that may never be fully overcome. One is the basic trade-off between spatial, spectral, and temporal resolution that is governed by basic laws of physics. High-resolution sensors of less than 5 m typically have narrow fields of view with small footprints and long repeat times, making them less useful for broadscale analyses. Though this could be considered a constraint, fortunately most environmental applications do not require such high-resolution imagery.

Another basic fact of life is that the data from passive sensors measure at sensor radiance, not what is actually on the ground. If space-based passive sensors were able to accurately, precisely, and repeatedly capture the actual reflectance from a feature on the ground, regardless of the time of day, season, or weather conditions, much of the hard work of image processing would be eliminated. But the reality is that the atmosphere scatters radiation that is reflected back out to space. Smoke, haze, clouds, and humidity exacerbate the problem and can block reflected energy entirely. This leads to gaps in the archival record. Radar or so-called “active” sensors can see through clouds, but the data are often difficult to interpret given the complicated reflectance patterns of different surface types. Where change detection is needed, there are further problems in discriminating actual on-the-ground changes from apparent changes due to illumination, view angle, or seasonality that affects ground cover.

Another issue is the fact that many remote sensing instruments are experimental, not operational, and this impacts the continuity of the data streams. Many of the most widely used sensors, such as MODIS or SPOT VEGETATION, are from experimental programs and the sensor characteristics will not be replicated in future missions. Even some operational sensors such as Landsat, which was until recently the most widely used sensor for environmental applications, have had data gaps due to technical failures and delays in planning continuity missions.

Beyond the constantly varying data streams, as new experimental sensors come online or older sensors fail, standard algorithms for many environmental applications have yet to be developed. The conclusion of one survey of remote sensing methods applied to biodiversity assessment and conservation found that there is great promise for their application, especially as new remote sensing instruments become available but that the field lacks a set of standard methodologies that could move

remote sensing applications from an experimental to an operational stage of implementation (de Sherbinin, 2005). In the realm of international law, replicability of observations and the reliability and consistency of measurements are of paramount importance (GOFC-GOLD, 2012). Algorithms also need to be validated through field studies or “ground truthing” that can be many times more expensive than the initial image processing.

As mentioned in the previous section, training is still somewhat of a barrier to wider application of remote sensing in the treaty realm, and this affects two communities. The first is the science community, which needs to ensure that adequate training programs are developed at undergraduate and graduate levels that ensure that students are grounded in the basic science behind remote sensing so that scientific findings are robust. The second is the policy community. Without a basic literacy in the science behind remote sensing, many contracting parties and treaty secretariats lack the ability to critically evaluate scientific findings. This results in either the uncritical acceptance of findings which may in fact be wrong (which is probably the more likely scenario) or the discounting or rejection of findings because of a lack of understanding of what the imagery tells us. This suggests that there is a need for increased literacy among the policy community concerning the science upon which remote sensing is based so as to make informed decisions.

On the policy side, there are other constraints. One of the areas which remote sensing scientists have been keenest in promoting is the potential for remote sensing to be used for treaty monitoring and enforcement. Although the remote sensing does hold out the promise of more rigorous regimes based on real monitoring, the reality is more complex. Because most environmental treaties are considered “soft law” based on consensus approaches to environmental management, treaty secretariats and Parties have typically shied away from even suggesting the use of remote sensing in a compliance context. Until more environmental treaties are crafted in such a way as to have enforceable requirements, the full promise of monitoring and enforcement applications will not be met.

Still, there have been research projects exploring the potential of remote sensing for compliance enforcement among the few agreements with requirements. One was in the case of the Bonn Agreement, officially known as The Agreement for Cooperation in Dealing with Pollution of the North Sea by Oil and Other Harmful Substances (1983). In that case, researchers were exploring the use of SAR images to detect oil slicks on the high seas (Jones, 2001). It was found that SAR gave an unacceptably high level of false-positives, and even if it could identify oil slicks, establishing the connection between a slick and a vessel would need to be corroborated by photographic evidence according to the terms of the convention. Another research project in Bolivia sought to apply a combination of ground-based measurements and aerial

videography to afforestation, reforestation, and deforestation (ARD) monitoring in the context of the Kyoto Protocol (see Noel Kempff Mercado Climate Action Project). Ultimately, to be applicable at broader scales, ARD monitoring will require a wider array of satellite sensors (including LIDAR) and more automated techniques than are currently available (Rosenqvist et al., 1999; CIESIN, 2001; GOFC-GOLD, 2012).

Other policy constraints include a lack of expertise on the part of treaty secretariats to assimilate remote sensing imagery or its results in their standard practices, and a fear by some, particularly developing country parties, of a “big brother” approach to compliance monitoring. Related to the former, it could be said that until recently there was a “technology push” on the part of space agencies and scientists rather than a genuine acceptance of remote sensing in the treaty world, lending the appearance of data developers in search of applications. An example was the European Space Agency’s Treaty Enforcement Services using Earth Observation (TESEO) project, later subsumed under the Data User Element for MEAs. This technology push, however, was not necessarily a bad thing. TESEO was successful in engaging a number of treaty secretariats such as that of the Ramsar Convention on Wetlands and the Convention to Combat Desertification (CCD) and informing their staff. One impact metric is that references to remote sensing or satellite imagery in Ramsar Conference of Party (COP) decision documents increased from no references in all up until 1999 and then three references at COP7 (1999), three references at COP8 (2002), and four references at COP9 (2005). In the case of UNCCD, there were no references to either term until COP8 (2007) (SEDAC, 2010).

For its part, the US National Aeronautics and Space Administration (NASA) engaged the Convention on Biological Diversity (CBD) secretariat as well as a number of partners through its NASA-NGO Biodiversity Working Group. Working with the CBD and the World Conservation Monitoring Centre, the working group published the *Sourcebook on Remote Sensing and Biodiversity Indicators* (Strand et al., 2007), which provides a good balance of technical and applied information concerning the application of remote sensing that is useful to the CBD community.

Conclusion

In conclusion, although a number of technical and political obstacles remain, in all likelihood remote sensing will play an increasing role in international environmental politics, including in the treaty realm. Treaty-specific applications are likely to be enhanced as treaty secretariats or those familiar with treaty provisions actually participate in instrument design teams. The GEOSS 10 year action plan, established in 2005, makes reference to a number of application areas that are central to environmental treaties such as the UN Framework Convention on Climate Change, the CBD, the CCD, and Ramsar, including:

- Improving management of energy resources
- Understanding, assessing, predicting, mitigating, and adapting to climate variability and change
- Improving water resource management through better understanding of the water cycle
- Improving the management and protection of terrestrial, coastal, and marine ecosystems
- Supporting sustainable agriculture and combating desertification
- Understanding, monitoring, and conserving biodiversity

The document specifically states that GEOSS will “further the implementation of international environmental treaty obligations.” An early workshop for the GEO Biodiversity Observation Network (Potsdam, Germany, April 2008) included significant representation of remote sensing scientists, and treaty secretariats have been engaged in the Network’s development.

Although remote sensing is not a panacea for the data and information needs of environmental treaties, and though even the best data does not guarantee action on the growing list of environmental problems, the data do represent a significant resource that will undoubtedly contribute to environmental assessments and serve to strengthen treaty mechanisms well into the future.

Bibliography

- American Museum of Natural History (AMNH), undated. *Biodiversity Informatics*. Available from World Wide Web: <http://biodiversityinformatics.amnh.org/>. Accessed June 24, 2008.
- SEDAC (Socioeconomic Data and Applications Center), 2010. *Environmental Treaty and Resource Indicators (ENTRI) COP Decision Search Tool*. Available from World Wide Web: http://sedac.ciesin.columbia.edu/gsametasearch/cop_start.jsp. Accessed September 12, 2010.
- Center for International Earth Science Information Network (CIESIN), Columbia University. 2001. *Remote Sensing and Environmental Treaties: Building More Effective Linkages*. Report of a workshop held 4–5 December 2000 at the Woodrow Wilson Center, Washington, DC. Available from World Wide Web: <http://sedac.ciesin.columbia.edu/rs-treaties/>. Accessed July 2, 2008.
- de Sherbinin, A., 2005. *Remote Sensing in Support of Ecosystem Management Treaties and Transboundary Conservation*. Report prepared for the project on Remote Sensing Technologies for Ecosystem Management Treaties funded by the United States Department of State, Bureau of Oceans and International Environmental and Scientific Affairs initiatives program (OESI). Available from World Wide Web: http://sedac.ciesin.columbia.edu/rs-treaties/RS&EMTreaties_Nov05_screen.pdf. Accessed June 26, 2008.

- de Sherbinin, A., Kline, K., and Raustiala, K., 2002. Remote sensing data: valuable support for environmental treaties. *Environment*, **44**(1), 20–31.
- GOCF-GOLD, 2012. *A Sourcebook of Methods and Procedures for Monitoring and Reporting Anthropogenic Greenhouse Gas Emissions and Removals Caused by Deforestation, Gains and Losses of Carbon Stocks in Forests Remaining Forests, and Forestation*. Available at http://www.gocfgold.wur.nl/redd/sourcebook/GOCF-GOLD_Sourcebook.pdf. Accessed June 19, 2012.
- Johnston, S., 2006. *Space in Environmental Diplomacy: Exploring the Role of Earth Observing Satellites for Monitoring International Environmental Agreements*. PhD thesis, The George Washington University. Available from <http://search.proquest.com/docview/305336614/abstract>. Accessed June 7, 2013.
- Jones, B., 2001. A comparison of visual observations of surface oil with synthetic aperture radar imagery of the sea empress oil spill. *International Journal of Remote Sensing*, **22**(9), 1619–1638.
- Mitchell, R., 2008. *International Environmental Agreements Database Project (Version 2007.1)*. Available from World Wide Web: <http://iea.uoregon.edu/>. Accessed July 2, 2008.
- Noel Kempff Mercado Climate Action Project web site. Available from World Wide Web: <http://www.forestcarbonportal.com/project/noel-kempff-mercado-climate-action-project>. Accessed June 7, 2013.
- Rosenqvist, A., Imhoff, M., Milne, A., and Dobson, C., 1999. *Remote Sensing and the Kyoto Protocol: A Review of Available and Future Technology for Monitoring Treaty Compliance*. Report of an ISPRS workshop, October 20–22, 1999, Ann Arbor, MI. Available from http://sedac.ciesin.columbia.edu/rs-treaties/Remote_Sensing_Kyoto.pdf. Accessed June 7, 2013.
- Strand, H., Hoft, R., Stritholt, J., Miles, L., Horning, N., and Fosnight, E. (eds.), 2007. *Sourcebook on Remote Sensing and Biodiversity Indicators*. CBD Technical Series No. 32. Montreal: Convention on Biological Diversity. Available from World Wide Web: <http://www.cbd.int/doc/publications/cbd-ts-32.pdf>. Accessed June 26, 2008.
- United Nations Environment Programme (UNEP), 2005. *One Planet Many People: Atlas of Our Changing Environment*. Nairobi: UNEP/UNEP Division of Early Warning and Assessment.

Cross-references

Air Pollution
 Atmospheric General Circulation Models
 Climate Monitoring and Prediction
 Coastal Ecosystems
 Fisheries
 Global Climate Observing System
 Global Earth Observation System of Systems (GEOSS)
 Global Land Observing System
 Rainfall
 Rangelands and Grazing
 Water Resources
 Wetlands

F

FIELDS AND RADIATION

Frank S. Marzano
Department of Information Engineering, Sapienza
University of Rome, Rome, Italy
Centre of Excellence CETEMPS, University of L'Aquila,
L'Aquila, Italy

Definition

Remote sensing is the technique to retrieve information about an object without being in physical contact with it (Elachi, 1987). This information is acquired by detecting and measuring changes that the object under investigation imposes on the surrounding field.

The interacting field can be an electromagnetic, acoustic, or gravitational field. The *electromagnetic field* is due to space–time variations of the electric and the magnetic fields (Ulaby et al., 1981; Ishimaru, 1991). Static electric and magnetic fields can be originated by a stationary electric charge distribution or by stationary currents or ferromagnetic materials, respectively. The electromagnetic methods, which are the most frequently used in remote sensing, cover the whole electromagnetic spectrum from radio-frequency waves to gamma rays through microwaves, submillimeter waves, and far infrared, near infrared, visible, ultraviolet, and x-ray waves. *Acoustic fields* are due to the continuous exchange between the fluid kinetic energy and the potential energy store during fluid compression (Lighthill, 1978). The acoustic refraction of Earth atmosphere depends on temperature, wind, and, to a lesser extent, on humidity. *Gravitational fields* are due to gravitational forces, typically exerted by a planetary mass. The gravitational fields determine the dynamics of objects embedded in the field itself.

Radiation represents the field propagation through waves. Amplitude, phase, polarization, and power are typical features of a field radiation (Elachi, 1987). The latter

can be generated by transformation of energy from other sources such as kinetic, chemical, thermal, electrical, magnetic, or nuclear. Several transformation mechanisms can lead to field waves over different regions of the frequency spectrum. Generally speaking, the more organized the transformation mechanism is, the more coherent is the generated radiation. From a theoretical point of view, radiation can be stated through the wave equation which describes the wave propagation phenomenon for electromagnetic, acoustic, and gravitational fields (Ishimaru, 1991). Physical properties of the media where the field is generated, such as temperature, humidity, shape, and composition, determine the medium properties which govern the field–matter interaction. The refractive index of the atmosphere is a typical example of electric properties which influence electromagnetic propagation.

Due to their inherent remote operation, remote sensing systems must exploit a *propagation mechanism*. The capability to infer the characteristics of the object under remote observation are derived from the properties of the wave field that has interacted, through emission, absorption, scattering, reflection, or transmission, with the object itself (Ishimaru, 1981). The retrieved field interactions are mainly due to reflection, scattering, and transmission in case of active remote sensing systems, whereas to absorption and emission in case of passive systems. The reflected or scattered radiation has generally the same frequency of the incident one; however, there exist processes which can produce irradiation at frequencies different from the incident wave one, such as Raman scattering and fluorescence, in case of electromagnetic waves. Transmission properties of media determine the capability of radiation to penetrate the medium itself.

Bibliography

Elachi, C., 1987. *Introduction to the Physics and Technology of Remote Sensing*. New York: Wiley.

- Ishimaru, A., 1991. *Electromagnetic Wave Propagation, Radiation and Scattering*. Englewood Cliffs, NJ: Academic/Prentice Hall.
- Lighthill, J., 1978. *Waves in Fluids*. Cambridge: Cambridge University Press.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing: Fundamentals and Radiometry*. Reading: Addison-Wesley, Vol. I.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)
[Media, Electromagnetic Characteristics](#)
[Radiation, Multiple Scattering](#)
[Radiation, Volume Scattering](#)
[Radiative Transfer, Theory](#)

FISHERIES

Cara Wilson
 Southwest Fisheries Science Center, NOAA/NMFS,
 Environmental Research Division, Pacific Grove,
 CA, USA

Definition

Catch per unit effort. CPUE standardizes fish catch data based on the amount of the effort (total time or area sampled) exerted.

Fishery stock. A subpopulation of a particular species in a given area. Unlike a fish population, a stock is defined as much by management concerns (such as jurisdictional boundaries or harvesting location) as by biology. There are three different types of stocks: commercial (exploited) species, unexploited species, and protected species. Stocks are not restricted to just fish, but also include marine mammals and invertebrates.

Fisheries stock assessment. An estimate of either the total population or total biomass of a fisheries stock. Stock assessments are a crucial component of fisheries management.

Fisheries management. The protection and management of fishery stocks to maintain sustainable exploitation for commercial species, and to recover the populations of protected and endangered species.

Fisheries oceanography. The study of oceanic processes affecting marine ecosystems and the relationship of these ecosystems to the abundance and availability of fish (Harrison and Parsons, 2001).

Living marine resources (LMR). A term to refer to all types of stocks, which is not restricted to just fish, but also includes marine mammals and invertebrates.

Operational fisheries. Utilizing oceanographic information to maximize the efficiency of fishing efforts.

Recruitment. The amount of fish added to the stock each year due to reproduction and/or migration into the stock area.

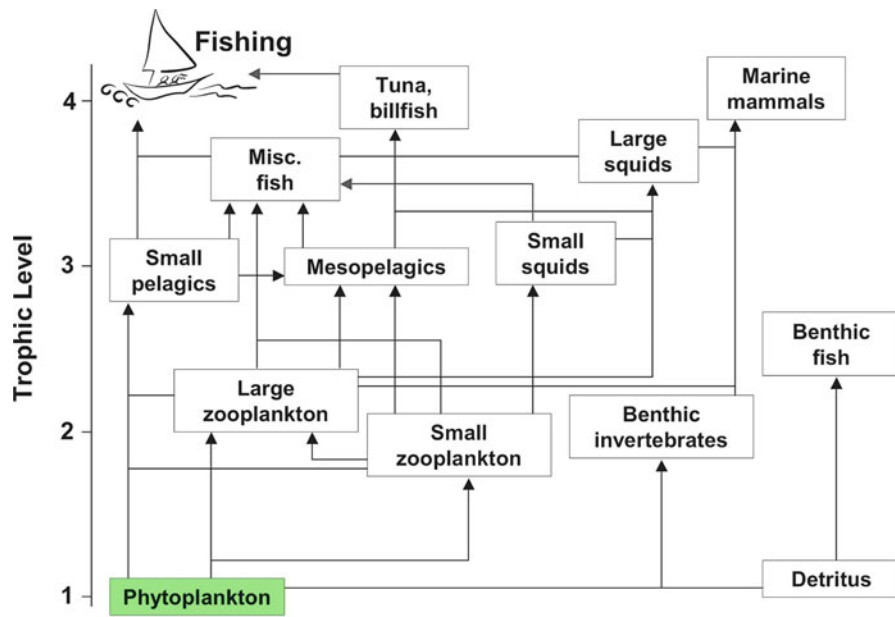
Fisheries and remote sensing

Introduction

In the broadest sense, fisheries encompass not just commercial fish stocks, but all living marine resources (LMRs), which for threatened and endangered species involve efforts to help the populations recover. There are three distinct aspects of fisheries: harvesting, assessment, and management, all of which have different goals. Harvesting efforts focus on increasing the catch per unit effort (CPUE), that is methods of finding, and more efficiently catching, more fish. Assessment involves both the species and its habitat. Stock assessments estimate either the total population or the total biomass of a fisheries stock in a given region, whereas habitat assessments characterize the environmental conditions favored by a species. Fisheries management uses both stock assessments and habitat assessments to set harvesting limits and guidelines on commercial stocks to maintain sustainable exploitation, and also to develop regulations to help recover the populations of protected and endangered species.

In the last half century, the world fish harvest has increased more than fourfold from 20 million tons in 1950 to over 90 million tons in 2000 (FAO Fisheries Department, 2004). At the same time the number of overexploited and depleted stocks has increased, and an expanding human population and problems with food supply have increased the pressure on fisheries resources. Better management and understanding of fisheries are needed to both maximize the utility of the current resources, and to ensure their sustainability into the future. However, these issues are complicated by the significant interannual variations that occur in fish populations, and sorting out fluctuations caused by anthropogenic effects (overexploitation, habitat alteration, pollution, etc.) from those caused by natural environmental variability is not trivial. The fundamental question of what drives the interannual variability of fish stocks was first posed over 100 years ago with the formation of the International Council for the Exploration of the Seas (ICES) in 1902, and still has not been adequately resolved (Kendall and Duker, 1998; Bakun and Broad, 2003).

Satellite data provide an environmental context within which to examine these issues, by measuring parameters of the habitat and ecosystems that influence marine resources at high temporal and spatial scales. There are two primary ways that satellite data are used within fisheries. One is to find populations, usually a commercial fish stock to increase CPUE, but also in some cases for conservation, for example, trying to identify locations of endangered cetaceans in order to minimize the number of lethal interactions with ships. A second application is characterizing and monitoring the habitat that influences living marine resources (LMRs). Most of the dynamic features that are important to ecosystems, that is, oceanfronts, eddies, convergence zones, river plumes, and coastal regions, cannot be adequately resolved without satellite data. Similarly,



Fisheries, Figure 1 Schematic representation of the ocean food web.

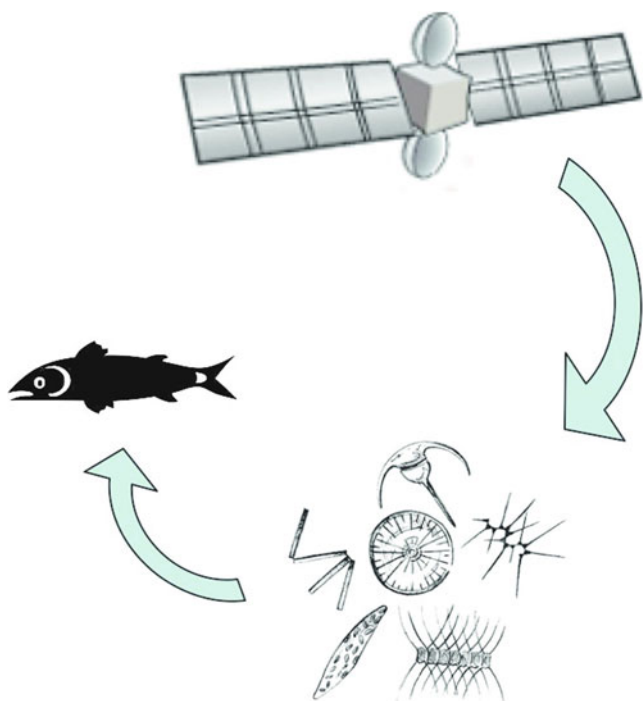
satellite data are crucial for resolving the timing of processes such as upwelling, harmful algal blooms, seasonal transitions, and El Niño events. Remotely sensed variables such as sea surface temperature (SST), sea surface height, ocean color, ocean winds, and sea ice are all used to characterize relevant environmental conditions (SSH, ocean color, ocean winds and sea ice, here and elsewhere). Additionally, environmental satellite data are used to monitor a number of issues that impact fisheries, such as coastal pollution and coral reef bleaching events. For some coastal applications, particularly aquaculture, high-resolution imagery data has been a useful tool. Many of these applications are described in more detail in the IOCCG's (International Ocean Colour Coordinating Group) report on remote sensing in fisheries and aquaculture (IOCCG, 2009) and in the proceedings of an international symposium on remote sensing and fisheries held in 2010 (Stuart et al., 2011).

Satellite ocean color is particularly important to fisheries, since it is the only remotely sensed parameter that directly measures a biological component of the ecosystem (Wilson et al., 2008). Satellite chlorophyll provides an index of phytoplankton biomass, which is the base of the oceanic food chain, or food web, as depicted in simplified form in Figure 1. The relationship between satellite chlorophyll data and a specific fish stock depends upon the number of linkages between phytoplankton and the higher trophic level. For some species, such as anchovies and sardines, which eat phytoplankton at some points in their life cycle, the linkage can be direct (Ware and Thomson, 2005), whereas for other species there are many trophic levels in between and the relationship can be nonlinear. There can also be spatial disconnects between satellite measurements of the ocean surface and demersal

and deepwater species. Nonetheless, chlorophyll is the only biological component of the marine ecosystem accessible to remote sensing, and as such it provides a key metric to measuring ecosystems on a global scale. Satellite chlorophyll measurements are the primary component in algorithms to calculate the primary productivity (PP) of the ocean. Global PP measurements, in conjunction with fish catch statistics and food web models, such as shown in Figure 2, can be used to estimate the carrying capacity of the world's fisheries. In the open ocean 2 % of the PP is needed to support the fishery catch, but in coastal regions the requirement ranges from 24 % to 35 %, suggesting that these systems are at or beyond their carrying capacity (Pauly and Christensen, 1995), which is a cause for concern as the bulk of the world's fish catch comes from coastal areas. In a similar manner, discrepancies between the values of satellite-derived PP and reported fish catches have been used to demonstrate spurious trends in global fish catches as reported by the Food and Agriculture Organization (FAO) of the United Nations (Watson and Pauly, 2001). In this instance satellite ocean color data provide an important objective baseline against which to gauge data that can have socioeconomic biases.

Operational fisheries (harvesting)

Locating and catching fish is becoming more challenging as fish stocks dwindle and move further offshore, thus increasing the search time, cost, and effort. Satellite data can help to increase CPUE by identifying oceanographic features that are often the sites of fish stock congregation and migration such as temperature fronts, meanders,



Fisheries, Figure 2 Cartoon of the connection between satellite measurements to fisheries through measurements of chlorophyll-bearing phytoplankton.

eddies, rings, and upwelling areas (Laurs et al., 1984; Fiedler and Bernard, 1987; Chen et al., 2005). Assuming adequate catch limitations are in place, increasing the CPUE is not incompatible with maintaining a sustainable fish stock population.

Both satellite ocean color and SST data have been used for increasing fishing efficiency. SST and ocean color often have similar patterns as generally warm, nutrient-depleted water has low chlorophyll concentrations and cold, nutrient-rich water has high chlorophyll. SST can also be an important factor for determining potential fishing grounds since different fish species have different optimal temperature ranges. Generally SST data have been used more often than ocean color data in fisheries applications. There are two main reasons for this. One, satellite SST is a more established data source, with data going back to 1981, whereas SeaWiFS, the first satellite to consistently provide satellite chlorophyll data on a global basis, was not launched until 1997. Second, SeaWiFS was a privately owned satellite, and while its data was always freely available, with a time delay, to the research community, the availability of near real-time (NRT) data needed by fishers for identifying potential fishing areas was only available on a commercial basis. For economic reasons that resulted in SeaWiFS data being unavailable to many fishers in the world, as the commercial costs of the real-time data can be prohibitive, particularly for those in underdeveloped countries, which is

where 70 % of the fish for human consumption comes from (FAO Fisheries Department, 2004). Ocean color data from the MODIS instrument on Aqua, launched in 2002, is available at no cost on an NRT basis.

For use in increasing the CPUE, satellite data must be available in a near real-time basis. There are international differences in how satellite data are disseminated to fishers, as national agencies serve different constituencies (Wilson, 2011). The mandate of NOAA Fisheries in the USA, for example, is to manage and conserve marine resources, and they are not allowed to provide services such as distributing “fish finding maps” that would compete with commercial interests. However in other countries, notably Japan and India, the national fisheries agencies are actively involved with helping increase the efficiency of their fishing fleets. For example, data from the Indian Ocean color satellite (in conjunction with satellite SST) are used by the Indian National Center for Ocean Information Services (INCOIS) to make maps of potential fishing zones (PFZ), which are freely disseminated several times a week throughout coastal India by fax, phone, Internet, electronic display boards, newspaper, and radio broadcasts. Studies on the effectiveness of the PFZ advisories have suggested that they have helped reduce search time by up to 70 % and have significantly increased the CPUE (Solanki et al., 2003; Zainuddin et al., 2004).

Stock assessment

While satellite remotely sensed data are now widely used in operational fisheries, their use in stock assessments is just beginning (Koeller et al., 2009). The incorporation of environmental data of any kind into stock assessment models has rarely been achieved successfully, for several reasons. Assessments have traditionally taken classical single species approaches which deal only with the numeric population dynamics of the stock under review – the fish stock as “bank account,” with principal (abundance), interest (growth), deposits (recruitment), and withdrawals (natural and fishing mortality) determined by research vessel surveys and fishery (catch, effort) data. The environmental factors forcing change to the bank account are complex, poorly understood, and difficult to measure; consequently, they have largely been excluded from traditional assessment models, greatly limiting their accuracy and effectiveness (Koeller et al., 2009). Additionally, radical changes to methodology, such as incorporating environmental data, would compromise the interannual time series of a stock population derived from stock assessments.

However, the advent of the “ecosystem approach to fisheries” (EAF) has given new impetus to better understand the environmental factors influencing fish stock dynamics and to try to include environmental variability as an integral part of the assessment process. It is particularly important to develop an understanding of the factors determining recruitment of commercially important fish and

shellfish stocks, for two reasons: first, the adverse effects of fishing cannot be separated from “normal” environmentally driven changes unless the latter are thoroughly understood. Second, environmental factors modify underlying stock-recruitment relationships, arguably the most important information necessary to define reference points and achieve fisheries sustainability. Until recently, defining stock-recruitment relationships and identifying the environmental factors modifying them have been the “holy grail” of fisheries research, largely unresolvable with traditional oceanographic methods because of the complex, large- and small-scale spatial/long- and short-term temporal processes involved. However, the availability of environmental satellite data such as ocean color, SST, and altimetry is now making these objectives achievable.

Recruitment

A fundamental issue in fisheries oceanography is understanding how environmental variability affects annual recruitment, the number of new individuals entering a stock. Recruitment is an important parameter because the bulk of mortality occurs in the development of larvae from eggs. Most fish have planktonic larval stages that are strongly influenced by ocean circulation and can have narrow ranges of optimal thermal conditions. Availability of a good food source is important for successful recruitment and hence many fish reproduce near the seasonal peak in phytoplankton abundance. A long-standing hypothesis in fisheries has been that recruitment success is tied to the degree of timing between spawning and the seasonal phytoplankton bloom, the Cushing-Hjort or match-mismatch hypothesis (Cushing, 1969, 1990). This hypothesis has been difficult to address with traditional shipboard measurements that have limited spatial and temporal resolution, but with satellite ocean color data, interannual fluctuations in the timing and extent of the seasonal bloom can be clearly seen. In an application on the Nova Scotia Shelf, the timing of the spring bloom determined from satellite ocean color was compared with available in situ data on larval survival of haddock, an important commercial fish species. Comparison of these two independent data sets indicated that highly successful year classes of haddock are associated with exceptionally early spring blooms of phytoplankton, confirming the match-mismatch hypothesis (Platt et al., 2003). A comparable study has also documented a relationship between the timing of the spring bloom and the growth rate of shrimp (Fuentes-Yaco et al., 2007). These studies demonstrate that it can be possible to separate ecosystem-associated variability in fish stocks from other components such as human exploitation or predation effects. The satellite time series permits the extraction of value-added products, in this case the timing of the seasonal biological cycle. Understanding these processes will lead to improved and longer term fisheries forecasts, that is, for the period between birth and capture, which for some species can be as much as a decade.

Habitat assessment

Coral reef monitoring

Coral reef ecosystems support a high diversity of coral, fish, and benthic species, with corals forming the structural and ecological foundation of the reef system. Coral reefs are sensitive to their environment (temperature, light, water quality, and hydrodynamics), and as a result of both anthropogenic and climate impacts (Kleypas et al., 2001), they are among the most threatened coastal ecosystems worldwide (Pandolfi et al., 2003; Hoegh-Guldberg et al., 2007). Corals have a symbiotic relationship with a microscopic organism, zooxanthellae, which provides the corals with oxygen and a portion of the organic compounds they produce through photosynthesis. When stressed, many reef inhabitants expel their zooxanthellae en masse. The polyps of the coral are left bereft of pigmentation and appear nearly transparent on the animal's white skeleton, a phenomenon referred to as coral bleaching.

Severe bleaching events can have dramatic long-term effects on the coral. Recovery rates appear to differ with species, and the time required to attain full recovery of symbiotic algae varies from as little as 2 months to as much as 1 year. When the level of environmental stress is high and sustained, the coral may die. Since the late 1980s, coral bleaching related to thermal stress has become more frequent and more severe. High SSTs associated with the 1997–1998 El Niño caused bleaching in much of the world's oceans, particularly in the Indian Ocean and in the western Pacific. Other major bleaching events occurred around the Great Barrier Reef and Northwestern Hawaiian Islands in 2002 and in the Caribbean in 2005.

With the capability of providing synoptic views of the global oceans in near real time and the ability to monitor remote reef areas, satellite remote sensing has become a key tool for coral reef managers and scientists (Mumby et al., 2004; Maina et al., 2008; Maynard et al., 2008). Since 1997, NOAA has been producing near-real-time, web-accessible, satellite-derived SST products to globally monitor conditions that might trigger coral bleaching from thermal stress. Currently NOAA's Coral Reef Watch Program provides operational products such as SST anomalies, bleaching hot spot anomalies, Degree Heating Weeks, and Tropical Ocean Coral Bleaching Indices to the global coral reef community (Strong et al., 2006). These products provide an effective early warning system globally, but are not always accurate in predicting the severity of a bleaching event at a regional scale (McClanahan et al., 2007; Maynard et al., 2008). In Australia, CSIRO's (Commonwealth Scientific and Industrial Research Organisation) ReefTemp project produces satellite-derived bleaching risk indices specifically for the Great Barrier Reef (Maynard et al., 2008).

Tagging

Electronic tagging of LMRs is a key methodology to gather information needed for accurate and responsible fisheries management. Satellite data is crucial to place

track data in an environmental context, in order to fully understand foraging and migration patterns, fish behavior and feeding ecology, habitat selection, and individual and population-level responses to environmental and climate variability. This approach has been used to characterize the environment of a wide variety of tagged species – turtles, penguins, seals, salmon, etc. – to better understand both their behavior and their habitat (Block et al., 2003; Hinke et al., 2005; Ream et al., 2005; Polovina et al., 2006; Weng et al., 2007). An example of this is described in fuller detail under the “Sea Turtles” subheading in the “Management of Protected Species” section.

Survey support

Fishery independent surveys are a crucial part of stock assessment. Just as satellite data can be used to increase fishing CPUE by identifying front locations and other features where fish tend to congregate, these data are also routinely used by fisheries cruises doing survey assessments for management and stock assessment. The near real-time data are valuable for locating fronts and other relevant features to sample across, as well as placing the results in a larger spatial context.

Fisheries management

Vessel monitoring system (VMS)

A standard fishery management tool is to close certain areas to fishing or to establish restricted fishing in certain areas, but ensuring that regulations and laws are being adhered to can be difficult for enforcement agencies. Satellite-based vessel monitoring systems (VMS) enable vast expanses of the ocean to be effectively monitored. By using a transmitter aboard commercial fishing vessels, paired with traditional global positioning satellites, vessels can be monitored while at sea to determine if they are fishing in closed areas or out of season. An obvious limitation of VMS is that vessels without installed VMS units, or vessels with faulty VMS units, are not monitored. In many fisheries, fishing vessels are required to carry an operational VMS transmitter, but illegal fishing is still possible by noncompliant fishing vessels. Recently synthetic aperture radar data have been used to successfully identify noncompliant fishing vessels, that is, vessels operating without VMS units (Kourti et al., 2005). In addition to their aid in enforcement of fisheries regulations, VMS data can provide valuable information to managers about both fish stock distributions and patterns of fishing activities (Deng et al., 2005; Mills et al., 2007; Bertrand et al., 2008).

Management of protected species

Monk seals

In the late 1980s, field programs monitoring monk seal pup survival, sea bird reproductive rates, and reef fish densities in the Northwestern Hawaiian Islands (NWHI) indicated ecosystem changes had occurred.

However, due to a lack of oceanographic data at sufficient space and time scales, it was difficult to construct environmental indicators, or envision how environmental variation might be coupled with the higher trophic level changes (Polovina et al., 1994). The launch of the SeaWiFS ocean color sensor in 1997 allowed assessment of basin-wide biological variability across the Pacific. The SeaWiFS imagery shows that during the winter, the northern atolls of the Hawaiian Archipelago, Kure, Midway, and Laysan Atolls are located at the boundary between the cool, high surface chlorophyll, vertically mixed water on the north and the warm, low surface chlorophyll, vertically stratified subtropical water on the south. This boundary has been termed the transition zone chlorophyll front (TZCF) (Polovina et al., 2001).

In some years, the TZCF remains north of these northern atolls throughout the year, while in other years, the TZCF shifts far enough south during the winter to encompass these atolls with higher chlorophyll water. Hence, the ecosystem of the northern atolls is more productive after the TZCF is located more southerly relative to its long-term winter position. Specifically during a winter when the TZCF was shifted south of its average position, monk seal pup survival increased 2 years later (Baker et al., 2007). The 2 year time lag probably represents the time needed for enhanced primary productivity to propagate up the food web to monk seal pup prey. Should management action, such as a head start program, be developed to improve pup survival, a 2 year forecast based on satellite ocean color can be used to predict the years when low survival is likely and hence when management intervention is needed.

Right whales

With fewer than 400 individuals left, the North Atlantic right whale is one of the most endangered whale populations (International Whaling Commission, 1998; Kraus et al., 2005). This population spends much of its time in US and Canadian waters, with the winter calving grounds off of Florida, Georgia, and South Carolina, and feeding grounds in the Gulf of Maine. The recovery of this population is limited by high mortality, especially due to ship strikes and entanglements in fishing gear. Because its habitat overlaps with lucrative fishing grounds and shipping lanes of major US ports, reducing mortality is politically and economically challenging (International Whaling Commission, 1998; Kraus et al., 2005). The current management strategy involves limiting adverse impacts by requiring modifications to fishing gear or vessel speeds in regions and time periods when whales are likely to be present. Thus, all management options require knowing when and where whales are likely to be. The question is how to identify these likely regions within a dynamic ocean environment.

A new approach to locating right whales combines synoptic information from satellites with a model of the right whales' main prey. Right whales feed on small crustaceans

called copepods, especially the large and abundant species *Calanus finmarchicus*. High numbers of whales are typically found in regions of high copepod concentrations (Pendleton et al., 2009). Many important rates in *Calanus*'s life cycle can be estimated using satellite data. The time required for an egg to develop into an adult is related to temperature, with shorter generation times in warmer water. Using satellite chlorophyll as a proxy for phytoplankton, the main food of *Calanus*, determines how quickly a female copepod can produce eggs. By combining the rate information derived from satellite data with reconstructions of the ocean currents from a computer model, estimated maps of *Calanus* abundance can be produced and related to right whale distributions (Pershing et al., 2009a, b). An initial test of this system forecasted that due to the cold winter in 2008, the *Calanus* population would be delayed, and that whales would arrive on their main spring feeding ground east of Cape Cod 3 weeks later than normal. While a full analysis of the data is underway, it appears that the whales arrived close to when the model predicted. These forecasts are currently being expanded to include a wider area of space and time and will soon be able to incorporate observations of both copepods and whales.

Sea turtles

A pelagic longline fishery based in Hawaii occasionally catches several species of sea turtles, with the threatened loggerhead sea turtle historically accounting for the majority of the turtle bycatch. Since 1997, Argos-linked transmitters have been attached to loggerhead sea turtles caught and released by longline vessels (Polovina et al., 2000), in order to characterize migration and forage areas of loggerheads, with the aim of spatially separating the fishery from the loggerheads. In recent years, the number of tracked turtles has been augmented by releasing hatchery reared loggerheads provided by the Port of Nagoya Aquarium, Nagoya, Japan. To characterize turtle habitat it is necessary to place their tracks within an environmental context. The use of satellite SST, ocean color, altimetry, and wind data have all been important in defining the oceanographic habitat of turtles within the North Pacific (Polovina et al., 2000, 2004, 2006; Kobayashi et al., 2008), allowing determination of seasonal habitat maps (Kobayashi et al., 2008). By combining this information with fisheries and fisheries bycatch data, it is now possible to predict the locations of areas with a high probability of loggerhead and longline interactions (Howell et al., 2008). In 2006, NOAA launched an experimental product called TurtleWatch, which uses satellite oceanographic data to map, in near real time, areas with a high probability of loggerhead and longline interactions, so that fishers can avoid them. This information benefits both the turtles and the fishers, who operate under strict limits on the number of turtle interactions allowed. The TurtleWatch tool is generated and distributed daily in near real time since the zone with the high probability of loggerhead bycatch is

a temporally dynamic feature. The TurtleWatch product is also provided to fishers onboard via a commercial fisheries information system.

Aquaculture

Many aquaculture species (e.g., bivalves, shrimp) are suspension feeders and derive nutrition from particulates, the most nutritious being phytoplankton. Bivalve aquaculture can have a significant impact on the local environment by reducing both the levels of particulates in the water, and of anthropogenic eutrophication (Lindahl et al., 2005). Given its ability to measure both phytoplankton and also turbidity (Hoepffner et al., 2008), which is detrimental for some species, satellite ocean color data can be a valuable tool in aquaculture development, but currently it is underutilized in this capacity (Grant et al., 2009). The primary reason for this underutilization is resolution – many aquaculture sites are below the spatial detection of single pixels, for example, small estuaries and bays 1 km wide (see *Coastal Ecosystems*). Various high-resolution commercial satellites have been used for aquaculture applications such as mapping mussel farms (Alexandridis et al., 2008), and site selection for nearshore aquaculture sites has been routinely based on the use of multispectral images from high spatial-resolution sensors (e.g., Landsat, Spot) which more recently have been complemented with the application of sensors such as Aster or IRS LISS/PAN (Dwivedi and Kandrika, 2005). However, medium resolution environmental satellite data can be used for open ocean applications. For example, site selection of sea bream and sea bass cages near the Canary Islands (Spain) made extensive use of SST data for the identification of suitable culture temperatures in the region (Pérez et al., 2003). Radar imagery (ERS-2 and RADARSAT) has been used to inventory and monitor milkfish cage culture in the Philippines (Travaglia et al., 2004).

Fisheries and climate

The SeaWiFS ocean color sensor was launched in August of 1997, just prior to the 1997/1998 El Niño which was one of the strongest ENSO (El Niño-Southern Oscillation) events of the century. This satellite data, in synergy with data from an extensive array of moorings across the equatorial Pacific, has contributed enormously to our understanding of ENSO dynamics and their ecosystem impacts. Deepening of the thermocline, and cessation of upwelling along the equator and in the coastal ecosystems, lowers ocean productivity and causes significant drops in the anchovy fisheries of Peru and Chile (Alamo and Bouchon, 1987; Escibano et al., 2004). However, other species are positively impacted by El Niño, for example, increases are observed in the biomass of sardine and mackerel (Bakun and Broad, 2003; Niquen and Bouchon, 2004). Satellite ocean color data have demonstrated that the effects of El Niño are not constrained to just the equatorial and coastal upwelling regions, but extend throughout most of the Pacific Ocean. For example,

during the 1997/1998 event the TZCF was shifted $\sim 5^\circ$ south of its regular position (Bograd et al., 2004), and lower chlorophyll values occurred across most of the subtropical Pacific (Wilson and Adamec, 2001).

Since we only have continuous ocean color data from 1997 onward, it is not possible to detect decade-scale variability with just these data (see *Climate Monitoring and Prediction*). However, it is possible to observe long-term changes by comparing climatological SeaWiFS data with data from the Coastal Zone Color Scanner (CZCS), which operated between 1979 and 1985. For example, the present wintertime position of the TZCF in the Pacific is about 5° further north than it was during CZCS time period, and this shift has also been seen in SST data used as a proxy for the TZCF (Bograd et al., 2004). Data from these two different satellites have also been used to demonstrate regions of the ocean which have experienced significant changes in the amount of chlorophyll and primary productivity in the past 20 years (Gregg and Conkright, 2002; Gregg et al., 2003).

There is significant long-term temporal variability in fish stocks, and for over 150 years, scientists have been trying to differentiate the effects of interannual variability, overfishing, and long-term changes such as regime shifts, which are characterized by relatively rapid changes in the baseline abundances of both exploited and unexploited species (Polovina, 2005). Long-term variations in ecosystems often follow trends or patterns also observed in ocean and atmosphere parameters (Mantua et al., 1997; Hare, and Mantua, 2000; Peterson and Schwing, 2003). For example, a shift in the North Pacific in the 1970s between a shrimp-dominated ecosystem to one populated primarily by several species of bottom-dwelling groundfish species coincided with a regional change from a cool to a warm climate (Botsford et al., 1997; Anderson and Piatt, 1999). While similar phenomena have been seen for many different stocks, and in all ocean basins, the mechanisms that link large-scale ocean and atmosphere dynamics to changes in population abundances are not always clear (Botsford et al., 1997; Baumann, 1998) and the relationships are not always constant over time (Solow, 2002). Ecosystem changes related to regime shifts are not in themselves harmful to the ecosystems as a whole (Bakun and Broad, 2003), but in order to maintain sustainability, management practices must be flexible enough to recognize and accommodate them (Polovina, 2005). One of the current limitations of satellite data is their relatively short time-spans. For fisheries applications it is crucial that climate quality records of ocean color be maintained so that existing satellite records will be able to serve as a benchmark against which to gauge future changes and to track historical variations.

Summary

Satellite data characterize oceanic properties of habitat and ecosystems that influence living marine resources at spatial and temporal resolutions that are impossible to achieve any other way. The high spatial resolution

provides an important geographical context for interpreting other data. The daily-to-weekly temporal resolution allows for effective monitoring of many oceanic features and permits the extraction of value-added products such as the timing of seasonal events. For fisheries applications it is crucial that climate quality records of ocean color be maintained so that existing satellite records will be able to serve as a benchmark against which to gauge future changes and to track historical variations. These time series of science-quality satellite data are needed to understand linkages between climate and ecosystems, and to characterize and monitor ecosystems as part of an ecosystem-based approach to fisheries management.

Bibliography

- Alamo, A., and Bouchon, M., 1987. Changes in the food and feeding of the sardine (*Sardinops sagax*) during the years 1980–1984 off the Peruvian coast. *Journal of Geophysical Research*, **92**, 14411–14415.
- Alexandridis, T. K., Topaloglou, C. A., Lazaridou, E., and Zalidis, G. C. 2008. The performance of satellite images in mapping aquacultures. *Ocean and Coastal Management*, **51**, 638–644.
- Anderson, P. J., and Piatt, J. F., 1999. Community reorganization in the Gulf of Alaska following ocean climate regime shift. *Marine Ecology Progress Series*, **189**, 117–123.
- Baker, J. D., Polovina, J. J., and Howell, E. A., 2007. Effect of variable oceanic productivity on the survival of an upper trophic predator, the Hawaiian monk seal *Monachus schauinslandi*. *Marine Ecology Progress Series*, **346**, 277–283.
- Bakun, A., and Broad, K., 2003. Environmental “loopholes” and fish population dynamics: comparative pattern recognition with focus on El Niño effects in the Pacific. *Fisheries Oceanography*, **12**, 458–473.
- Baumann, M., 1998. The fallacy of the missing middle: physics $\rightarrow$... $\rightarrow$ fisheries. *Fisheries Oceanography*, **7**, 63–65.
- Bertrand, S., Diaz, E., and Lengaigne, M., 2008. Patterns in the spatial distribution of Peruvian anchovy (*Engraulis ringens*) revealed by spatially explicit fishing data. *Progress in Oceanography*, **79**, 379–389.
- Block, B. A., Costa, D. P., Boehlert, G. W., and Kochevar, R. E., 2003. Revealing Pelagic habitat use: the tagging of Pacific Pelagics program. *Oceanologica Acta*, **25**, 255–266.
- Bograd, S., Foley, D. G., Schwing, F. B., Wilson, C., Polovina, J. J., and Howell, E. A., 2004. On the seasonal and interannual migrations of the transition zone chlorophyll front. *Geophysical Research Letters*, doi:10.1029/2004GL020637.
- Botsford, L. W., Castilla, J. C., and Peterson, C. H., 1997. The management of fisheries and marine ecosystems. *Science*, **277**, 509–515.
- Chen, I.-C., Lee, P.-F., and Tzeng, W.-N., 2005. Distribution of albacore (*Thunnus alalunga*) in the Indian Ocean and its relation to environmental factors. *Fisheries Oceanography*, **14**, 71–80.
- Cushing, D. H., 1969. The regularity of the spawning season of some fishes. *Journal of Physical Oceanography*, **33**, 81–92.
- Cushing, D. H., 1990. Plankton production and year-class strength in fish populations – an update of the match mismatch hypothesis. *Advances in Marine Biology*, **26**, 249–294.
- Deng, R., Dichmont, C., Milton, D., Haywood, M., Vance, D., Hall, N., and Die, D., 2005. Can vessel monitoring system data also be used to study trawling intensity and population depletion? The example of Australia’s northern prawn fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 611–622.

- Dwivedi, R. S., and Kandrika, S., 2005. Delineation and monitoring of aquaculture areas using multitemporal space-borne multi-spectral data. *Current Science*, **89**, 1414–1421.
- Escribano, R., Daneri, G., Fariás, L., Gallardo, V. A., González, H. E., Gutiérrez, D., Lange, C. B., Morales, C. E., Pizarro, O., Ulloa, O., and Braun, M., 2004. Biological and chemical consequences of the 1997–1998 El Niño in the Chilean coastal upwelling system: a synthesis. *Deep-Sea Research, II*, **51**, 2389–2411.
- FAO Fisheries Department, 2004. *The State of World Fisheries and Aquaculture (SOFIA)*. Rome: U.N. Food and Agriculture Organization.
- Fiedler, P. C., and Bernard, H. J., 1987. Tuna aggregation and feeding near fronts observed in satellite imagery. *Continental Shelf Research*, **7**, 871–881.
- Fuentes-Yaco, C., Koeller, P. A., Sathyendranath, S., and Platt, T., 2007. Shrimp (*Pandalus borealis*) growth and timing of the spring phytoplankton bloom on the Newfoundland-Labrador Shelf. *Fisheries Oceanography*, **16**, 116–129.
- Grant, J., Bacher, J., Ferreira, J., Groom, S., Morales, J., Rodriguez-Benito, C., Saitoh, S.-i., Sathyendranath, S., and Stuart, V., 2009. Remote sensing applications to marine aquaculture. In Forget, M.-H., Stuart, V., and Platt, T. (eds.), *Remote Sensing in Fisheries and Aquaculture*. Dartmouth: IOCCG, pp. 77–87.
- Gregg, W. W., and Conkright, M. E., 2002. Decadal changes in global ocean chlorophyll. *Geophysical Research Letters*, **29**, 1730–1734, doi:10.1029/2002GL014689.
- Gregg, W. W., Conkright, M. E., Ginoux, P., O'Reilly, J. E., and Casey, N. W., 2003. Ocean primary production and climate: global decadal changes. *Geophysical Research Letters*, **30**, 1809, doi:10.1029/2003GL016889.
- Hare, S. R., and Mantua, N. J., 2000. Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Progress in Oceanography*, **47**, 103–145.
- Harrison, P. J., and Parsons, T. R. (eds.), 2001. *Fisheries Oceanography: an Integrative Approach to Fisheries Ecology and Management*. Oxford: Blackwell Science.
- Hinke, J. T., Foley, D. G., Wilson, C., and Watters, G. M., 2005. Persistent habitat use by Chinook salmon (*Oncorhynchus tshawytscha*) in the coastal ocean. *Marine Ecology Progress Series*, **304**, 207–220.
- Hoegh-Guldberg, O., Mumby, P. J., Hooten, A. J., Steneck, R. S., Greenfield, P., Gomez, E., Harvell, C. D., Sale, P. F., Edwards, A. J., Caldeira, K., Knowlton, N., Eakin, C. M., Iglesias-Prieto, R., Muthiga, N., Bradbury, R. H., Dubi, A., and Hatzioolos, M. E., 2007. Coral reefs under rapid climate change and ocean acidification. *Science*, **318**, 1737–1742.
- Hoepffner, N., Brown, C., Doerffer, R., Ishizaka, J., Gower, J., and Lynch, M., 2008. Ocean colour radiometry and water quality. In Platt, T., Hoepffner, N., Stuart, V., and Brown, C. W. (eds.), *Why Ocean Colour? The Societal Benefits of Ocean-Colour Technology*. Dartmouth: IOCCG, pp. 59–73.
- Howell, E. A., Kobayashi, D. R., Parker, D. M., Balazs, G. H., and Polovina, J., 2008. TurtleWatch: a data product to aid in the bycatch reduction of loggerhead turtle (*Caretta caretta*) in the Hawaii-based pelagic longline fishery. *Endangered Species Research*, **5**, 267–278.
- International Whaling Commission, (1998). Report of the workshop on the comprehensive assessment of right whales: A worldwide comparison. SC/50/REP 4.
- IOCCG, 2009. *Remote Sensing in Fisheries and Aquaculture*. Dartmouth: IOCCG.
- Kendall, A. W., Jr., and Duker, G. J., 1998. The development of recruitment fisheries oceanography in the United States. *Fisheries Oceanography*, **7**, 69–88.
- Kleypas, J. A., Buddemeire, R. W., and Gattuso, J.-P., 2001. The future of coral reefs in an age of global change. *International Journal of Earth Sciences*, **90**, 426–437.
- Kobayashi, D. R., Polovina, J. J., Parker, D. M., Kamezaki, N., Cheng, I., Uchida, I., Dutton, P., and Balazs, G., 2008. Pelagic habitat characterization of loggerhead sea turtles, *Caretta caretta*, in the North Pacific Ocean (1997–2006): insights from satellite tag tracking and remotely sensed data. *Journal of Experimental Marine Biology and Ecology*, **356**, 96–114.
- Koeller, P., Friedland, K., Fuentes-Yaco, C., Han, G., Kulka, D., O'Reilly, J., Platt, T., Richards, A., and Taylor, M., 2009. Remote sensing applications in stock assessments. In Forget, M.-H., Stuart, V., and Platt, T. (eds.), *Remote Sensing in Fisheries and Aquaculture*. Dartmouth: IOCCG, pp. 29–42.
- Kourti, N., Shepherd, I., Greidanus, H., Alvarez, M., Aresu, E., Bauna, T., Chesworth, J., Lemoine, G., and Schwartz, G., 2005. Integrating remote sensing in fisheries control. *Fisheries Management and Ecology*, **12**, 295–307.
- Kraus, S. D., Brown, M. W., Caswell, H., Clark, C. W., Fujiwara, M., Hamilton, P. K., Kenney, R. D., Knowlton, A. R., Landry, S., Mayo, C. A., McLellan, W. A., Moore, M. J., Nowacek, D. P., Pabst, D. A., Read, A. J., and Rollan, R. M., 2005. North Atlantic right whales in crisis. *Science*, **309**, 561–562.
- Laur, R. M., Fiedler, P. C., and Montgomery, D. R., 1984. Albacore tuna catch distributions relative to environmental features observed from satellites. *Deep Sea Research*, **31**, 1085–1099.
- Lindahl, O., Hart, R., Hernroth, B., Kollberg, S., Loo, L.-O., Olrog, L., Rehnstam-Holm, A.-S., Svensson, J., Svensson, S., and Syversen, U., 2005. Improving marine water quality by mussel farming: a profitable solution for Swedish society. *Ambio*, **34**, 131–138.
- Maina, J., Venus, V., McClanahan, M. R., and Ateweberhan, M., 2008. Modelling susceptibility of coral reefs to environmental stress using remote sensing data and GIS models. *Ecological Modelling*, **212**, 180–199.
- Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., and Francis, R. C., 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the American Meteorological Society*, **78**, 1069–1079.
- Maynard, J. A., Turner, P. J., Anthony, K. R. N., Baird, A. H., Berkelmans, R., Eakin, C. M., Johnson, J., Marshall, P. A., Packer, G. R., Rea, A., and Willis, B. L., 2008. ReefTemp: an interactive monitoring system for coral bleaching using high-resolution SST and improved stress predictors. *Geophysical Research Letters*, **35**, L05603, doi:10.1029/2007GL032175.
- McClanahan, T. R., Ateweberhan, M., Sebastia, C. R., Graham, N. A. J., Wilson, S. K., Bruggemann, J. H., and Guillaume, M. M., 2007. Predictability of coral bleaching from synoptic satellite and in situ temperature observations. *Coral Reefs*, **26**, 695–701.
- Mills, C. M., Townsend, S. E., Jennings, S., Eastwood, P. D., and Houghton, C. A., 2007. Estimating high resolution trawl fishing effort from satellite-based vessel monitoring system data. *ICES Journal of Marine Science*, **64**, 248–255.
- Mumby, P. J., Skirving, W., Strong, A. E., Hardy, J. T., LeDrew, E. F., Hochberg, E. J., Stumpf, R. P., and David, L. T., 2004. Remote sensing of coral reefs and their physical environment. *Marine Pollution Bulletin*, **48**, 219–228.
- Ñiquen, M., and Bouchon, M., 2004. Impact of El Niño events on pelagic fisheries in Peruvian waters. *Deep-Sea Research, II*, **51**, 563–574.
- Pandolfi, J. M., Bradbury, R. H., Sala, E., Hughes, T., Bjørndal, K. A., Cooke, R., McArdle, D., McClenahan, L., Newman, M., Paredes, G., Warner, R., and Jackson, J., 2003. Global trajectories of the long-term decline of coral reef ecosystems. *Science*, **301**, 955–958.
- Pauly, D., and Christensen, V., 1995. Primary production required to sustain global fisheries. *Nature*, **374**, 255–257.
- Pendleton, D. E., Pershing, A. J., Brown, M. W., Mayo, C. A., Kenney, R. D., Record, N. R., and Cole, T. V. N., 2009.

- Regional-scale mean copepod concentration indicates relative abundance of North Atlantic right whales. *Marine Ecology Progress Series*, **378**, 211–225.
- Pérez, O. M., Ross, L. G., Telfer, T. C., and del Campo Barquin, L. M. 2003. Water quality requirements for marine fish cage site selection in Tenerife (Canary Islands): predictive modelling and analysis using GIS. *Aquaculture*, **224**, 51–69.
- Pershing, A. J., Record, N. R., Monger, B. C., Mayo, C. A., Brown, M. W., Cole, T. V. N., Kenney, R. D., Pendleton, D. E., and Woodard, L. A., 2009a. Model-based estimates of right whale habitat use in the Gulf of Maine. *Marine Ecology Progress Series*, **378**, 245–257.
- Pershing, A. J., Record, N. R., Monger, B. C., Pendleton, D. E., and Woodard, L. A., 2009b. Model-based estimates of *Calanus finmarchicus* abundance in the Gulf of Maine. *Marine Ecology Progress Series*, **378**, 227–243.
- Peterson, W. T., and Schwing, F. B., 2003. A new climate regime in Northeast Pacific ecosystems. *Geophysical Research Letters*, **30**, 1896, doi:1810.1029/2003GL017528.
- Platt, T., Fuentef-Yaco, C., and Frank, K. T., 2003. Spring algal bloom and larval fish survival. *Nature*, **423**, 398–399.
- Polovina, J. J., 2005. Climate variation, regime shifts, and implications for sustainable fisheries. *Bulletin of Marine Science*, **76**, 233–244.
- Polovina, J. J., Mitchum, G. T., Graham, N. E., Craig, M. P., DeMartini, E. E., and Flint, E. N., 1994. Physical and biological consequences of a climate event in the central North Pacific. *Fisheries Oceanography*, **3**, 15–21.
- Polovina, J. J., Kobayashi, D. R., Parker, D. R., Seki, M. P., and Balazs, G. H., 2000. Turtles on the edge: movement of loggerhead turtles (*Caretta caretta*) along oceanic fronts spanning longline fishing grounds in the central North Pacific, 1997–1998. *Fisheries Oceanography*, **9**, 71–82.
- Polovina, J. J., Howell, E., Kobayashi, D. R., and Seki, M. P., 2001. The transition zone chlorophyll front, a dynamic global feature defined migration and forage habitat for marine resources. *Progress in Oceanography*, **49**, 469–483.
- Polovina, J. J., Balazs, G. H., Howell, E. A., Parker, D. M., Seki, M. P., and Dutton, P. H., 2004. Forage and migration habitat of loggerhead (*Caretta caretta*) and olive ridley (*Lepidochelys olivacea*) sea turtles in the central North Pacific Ocean. *Fisheries Oceanography*, **13**, 36–51.
- Polovina, J. J., Uchida, I., Balazs, G., Howell, E. A., Parker, D., and Dutton, P., 2006. The Kuroshio Extension Bifurcation Region: a pelagic hotspot for juvenile loggerhead sea turtles. *Deep-Sea Research*, **53**, 326–339.
- Ream, R. R., Sterling, J. T., and Loughlin, T. R., 2005. Oceanographic features related to northern fur seal migratory movements. *Deep-Sea Research II*, **52**, 823–843.
- Solanki, H. U., Dwivedi, R. M., Nayak, S. R., Somvanshi, V. S., Gulati, D. K., and Pattnayak, S. K., 2003. Fishery forecast using OCM chlorophyll concentration and AVHRR SST: validation results off Gujarat coast, India. *International Journal of Remote Sensing*, **24**, 3691–3699.
- Solow, A. R., 2002. Fisheries recruitment and the North Atlantic oscillation. *Fisheries Research*, **54**, 295–297.
- Strong, A. E., Arzayus, F., Skirving, W., and Heron, S. F., 2006. Identifying coral bleaching remotely via Coral Reef Watch: improving integration and implications for changing climate. In Phinney, J. T. (ed.), *Coral Reefs and Climate Change: Science Management*. Washington DC: AGU, pp. 163–180.
- Stuart, V., Platt, T., Sathyendranath, S., and Pravin, P., 2011. Remote sensing and fisheries: an introduction. *ICES Journal of Marine Science*, **68**, 639–641.
- Travaglia, C., Profeti, G., Aguilar-Manjarrez, J., and López, N. A. 2004. Mapping coastal aquaculture and fisheries structures by satellite imaging radar. Case study of the Lingayen Gulf, the Philippines, FAO Fisheries Technical Papers Rep. 459, Rome.
- Ware, D. M., and Thomson, R. E., 2005. Bottom-up ecosystem trophic dynamics determine fish production in the Northeast Pacific. *Science*, **308**, 1280–1285.
- Watson, R., and Pauly, D., 2001. Systematic distortions in world fisheries catch trends. *Nature*, **414**, 534–536.
- Weng, K. C., Boustany, A. M., Pyle, P., Anderson, S. D., Brown, A., and Block, B. A., 2007. Migration and habitat of white sharks (*Carcharodon carcharias*) in the eastern Pacific Ocean. *Marine Biology*, **152**, 877–894.
- Wilson, C., 2011. The rocky road from research to operations for satellite ocean-colour data in fishery management. *ICES Journal of Marine Science*, **68**, 677–686.
- Wilson, C., and Adamec, D., 2001. Correlations between surface chlorophyll and sea surface height in the tropical Pacific during the 1997–1999 El Niño-Southern Oscillation event. *Journal of Geophysical Research*, **106**, 31175–31188.
- Wilson, C., Morales, J., Nayak, S., Asanuma, I., and Feldman, G., 2008. Ocean colour radiometry and fisheries. In Platt, T., Hoepffner, N., Stuart, V., and Brown, C. W. (eds.), *Why Ocean Colour? The Societal Benefits of Ocean-Colour Technology*. Dartmouth: IOCCG, pp. 47–58.
- Zainuddin, M., Saitoh, S.-i., and Saitoh, K., 2004. Detection of potential fishing ground for albacore tuna using synoptic measurements of ocean color and thermal remote sensing in the northwestern North Pacific. *Geophysical Research Letters*, **31**, L20311, doi:20310.21029/22004GL021000.

Cross-references

[Climate Monitoring and Prediction](#)
[Coastal Ecosystems](#)
[Ocean Measurements and Applications, Ocean Color](#)
[SAR-Based Bathymetry](#)
[Sea Ice Concentration and Extent](#)
[Sea Surface Temperature](#)
[Sea Surface Wind/Stress Vector](#)

FORESTRY

Dar Roberts

Department of Geography, University of California, Santa Barbara, CA, USA

Synonyms

Forest conservation; Forest management/forest ecosystem management; Non-timber forest products; Silviculture; Sustainable forest management; Timber management

Definition

Forestry. The science of planting, monitoring, describing, and managing forests and forest systems and tree plantations for their goods and services, often for commercial production, natural resource use, and habitat preservation. More recently, consideration of the role of forests in carbon cycling and cultural and spiritual values is included.

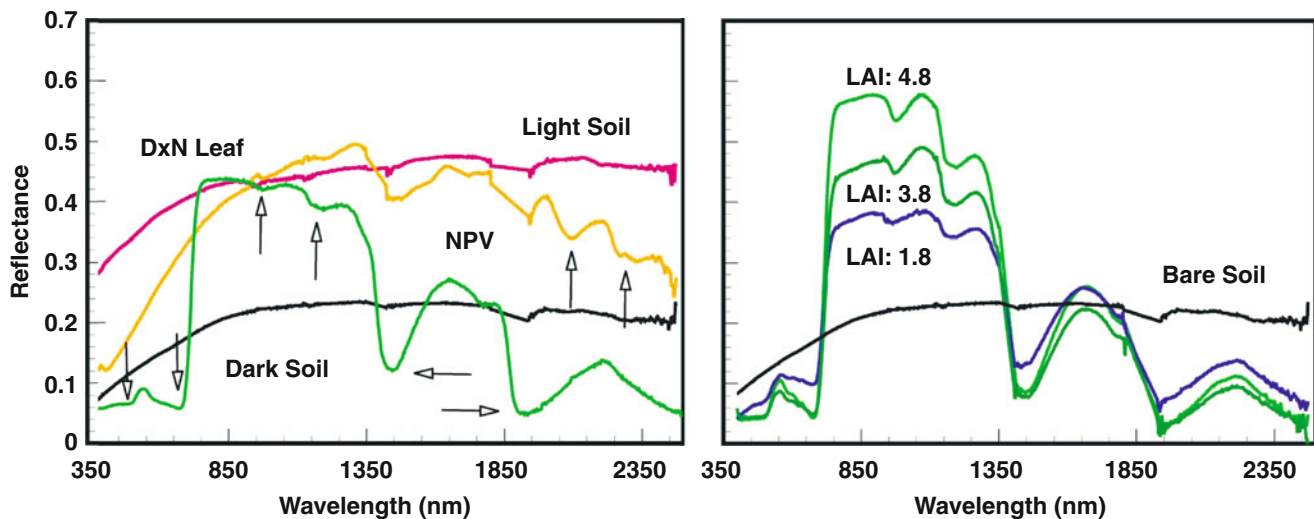
Introduction

Remote sensing has a long history in forestry, starting with the use of aerial photography in the early 1900s and progressing to more advanced satellite and airborne sensors today. While photointerpretation of aerial photography remains an important tool (Wulder, 1998), the launch of Landsat in 1972, and proliferation of numerous digital airborne and spaceborne sensors since then have broadened the types of research questions that can be asked, and the spatial area over which forests can be mapped and monitored.

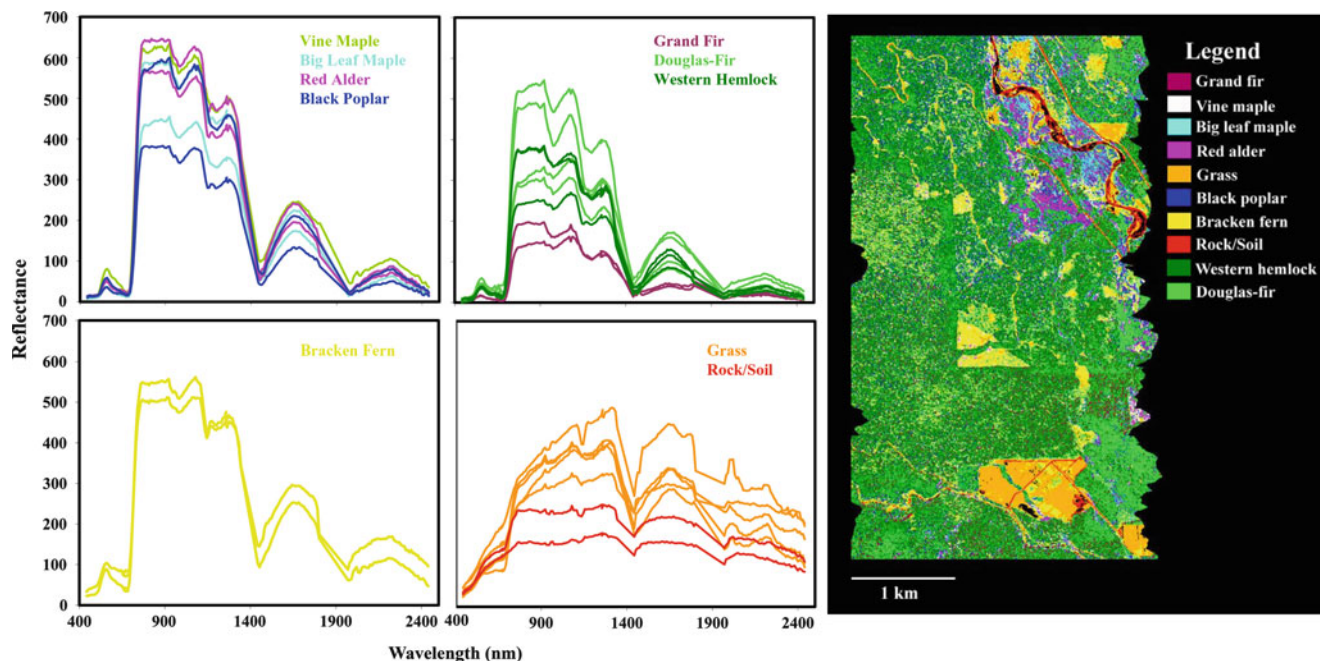
The wavelength-dependent intensity of electromagnetic radiation reflected or emitted from vegetation is a product of light scattering and absorption processes occurring at multiple scales. At the finest scales at or below the scale of an individual leaf or branch, the interaction varies depending on anatomy (i.e., thickness, internal structure, optical discontinuities between cell membranes) and biochemistry (molecular absorptions due to water, pigments, and other biochemical compounds and chlorophyll fluorescence; Gates et al., 1965). At canopy scales, the interaction depends upon the number of scatterers and absorbers and their physical arrangement, such as the number of leaves per unit volume and their angular distributions, the ratio of leaves to branches, crown geometry, and tree height. At stand scales, important stand attributes include tree density (trees/unit area), species composition, and stand area. In the reflected solar spectrum (0.350–2.5 μm), the manner in which these

multiscale processes modify reflected light can be used to infer important plant biochemical or structural properties, such as the number of leaves per unit ground area, known as Leaf Area Index (LAI: Figure 1). Biochemical, anatomical, and structural differences can also be used to discriminate tree species (Figure 2), map invasive tree species or forest pathogens (Wulder et al., 2006), and infer species diversity (Asner et al., 2009). In the Mid- and Thermal infrared (IR: 2.5–14 μm), emission depends on the temperature of individual canopy components (leaves and branches) and wavelength-dependent emission efficiency, called emissivity (Salisbury and Milton, 1988). Emissivity is very high for water-filled leaves and lower for branches and stems. Given knowledge of emissivity, canopy temperatures can be estimated and used to infer important plant physiological processes, such as rates of evapotranspiration. In the microwave (>1 mm), scattering and absorption depend on the physical size of canopy components, their orientation, and their dielectric properties, which are largely controlled by water content. Large, water-filled branches are highly reflective in the microwave.

Remote sensing systems can be broadly divided into active and passive sensors. Active remote sensing systems emit a directional energy pulse that interacts with a surface and then returns to the sensor. Information is determined about the surface based on the length of time it takes for the pulse to return, the strength of the returned pulse, and other attributes such as a change in polarization. The two



Forestry, Figure 1 Poplar leaf (left) and canopy reflectance (right). Leaf scale spectra are shown for a *Populus deltoides* x *Pinus nigra* hybrid and compared to two soils and Poplar wood (labeled NPV). Important biochemical absorptions are marked with arrows including chlorophyll (480 and 680 nm), water (980, 1,200, 1,450, and 1,900 nm), and lignocellulose (broad regions at 2,100 and 2,300 nm). Spectra on the right show how reflectance changes with increasing LAI, including an increase in NIR and decrease in red reflectance. Liquid water bands at 980 and 1,200 nm are enhanced as light encounters more leaves with an increase in LAI (Roberts et al., 2004) (Adapted from Davis and Roberts, 1999).



Forestry, Figure 2 Spectra of four broadleaf tree species, three conifers, bracken fern, senesced grass, and rock/soil, including multiple spectra of several of the species (*left*). Spectral differences within a plant species illustrate that spectra are often not unique, but vary with architecture, illumination, or other factors. Map of plant species at Wind River, Washington, generated using 2003 AVIRIS data (*right*).

primary active remote sensing systems used for forestry applications are radio detection and ranging (radar), which uses wavelengths of electromagnetic radiation longer than a millimeter, and light detection and ranging (lidar), which most often uses NIR lasers (0.9–1.064 μm). Active remote sensing is the primary tool used for mapping forest structure (e.g., tree height, crown properties, cover, and aboveground biomass). Passive remote sensing systems measure electromagnetic radiation that is either reflected solar radiation or emitted from the surface. Passive remote sensing systems include aerial photography and airborne or spaceborne scanners that measure a few to many wavelengths of light. Passive remote sensing systems play a greater role in mapping changes in forest cover, forest health, and plant chemistry. Both passive and active remote sensing methods are employed at the ground level, including the use of hemispherical photographs to estimate crown closure and LAI and laser range finders and ground-based lidar to estimate tree height (Davis and Roberts, 1999).

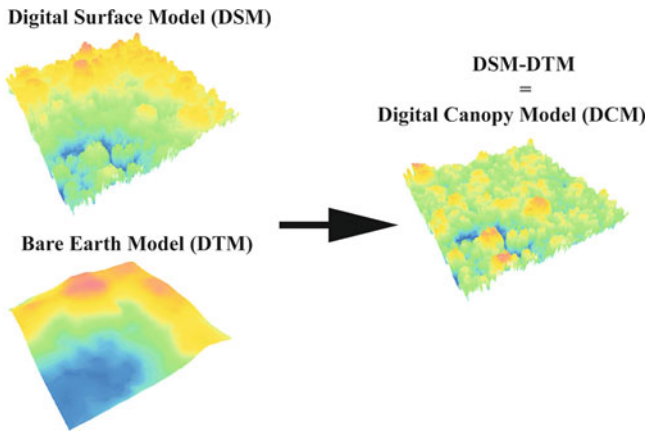
Remote sensing applications

Forest inventories

One of the primary remote sensing forestry applications has been to aid in the development of forest inventories. Forest inventories are designed to quantify important

properties of forest stands, including stand density (the number of trees per unit area), crown closure, average crown diameters, tree heights, and species composition (Avery and Burkhart, 1983). Species composition can be determined using dichotomous photointerpretation keys based on such measures as branching patterns, crown shape, the presence or absence of leaves, and color for color or color-infrared film (Heller and Ulliman, 1983). Tree height can be estimated using stereoscopic techniques, or geometrically from information on sun angles, viewing geometries, and shadows cast by trees. Stand volume and aboveground biomass can be estimated given accurate estimates of height, crown closure, and species composition using relationships derived from field methods. Aerial photographs also aid in improved field sampling by providing forest cover area and through improved stratified random sampling. Several national-scale inventory programs have relied extensively on stratified random sampling from aerial photography, including the forest inventory (Analysis Program: FIA: <http://fia.fs.fed.us/>) in the United States. Species-level differences in leaf optical and structural properties make it possible to discriminate individual tree species, using sensors such as the Advanced Visible/Infrared Imaging Spectrometer (AVIRIS: Roberts et al., 2004; Figure 2).

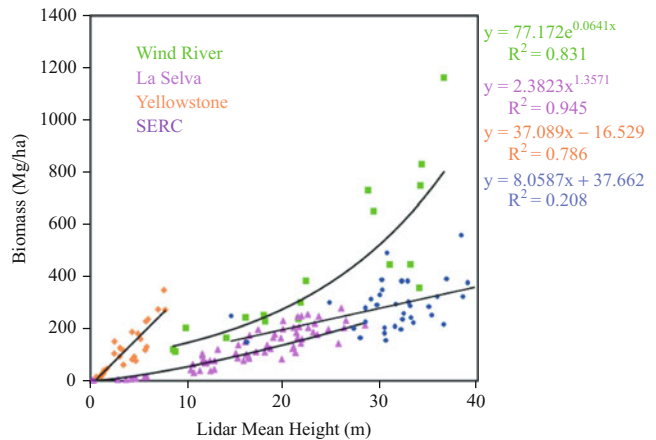
While coarser spatial resolution, spaceborne sensors can provide many important components of a forest



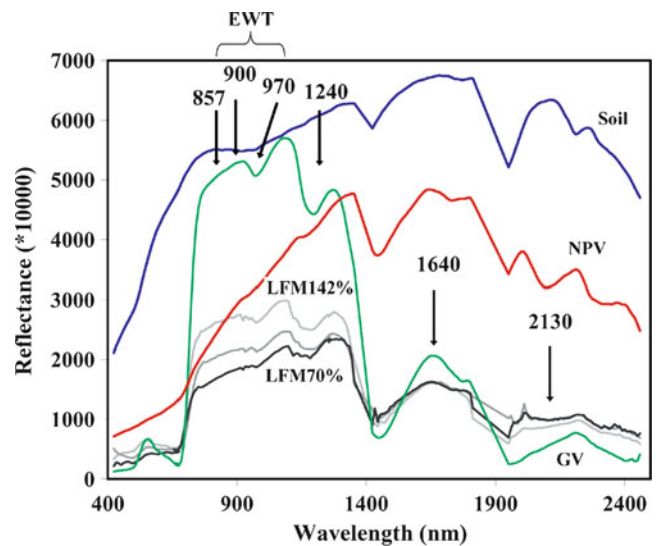
Forestry, Figure 3 Showing how a digital canopy model (DCM) is calculated from a digital surface model (DSM) and bare earth model (digital terrain model: DTM). This DSM was generated from a first return lidar system flown over La Selva, Costa Rica. The DTM was calculated using the same data by searching for ground returns within a fixed search window (Adapted from Clark et al., 2004).

inventory, the greatest advances have been made in the use of lidar. Lidar systems are typically classified as either discrete return or waveform (Lefsky et al., 2002). Discrete-return lidar systems typically illuminate a small area of ground and divide the returning laser pulse into one or more returns defined by major energy peaks in the returned signal (Lefsky et al., 2002). The returned signal typically includes a distance measure and the strength of return (intensity). Waveform lidar systems divide the returning pulse into a large number of predefined height bins, thus capturing changes in the intensity of reflected energy at a uniform vertical height interval. A majority of forestry applications rely on discrete-return lidar due to greater data availability and established utility.

Discrete-return lidar, by measuring vertical canopy structure, provides many of the structural elements critical for forest inventories. Examples include crown closure, calculated as the percentage of crown returns to total area imaged and canopy height, estimated as the difference between a ground return and the highest return for a specific pixel (digital canopy model, or DCM: Figure 3). Crown diameter can be determined provided individual crowns can be identified and tree height can be estimated from the highest return within an identified crown. Crown base height can be determined from lidar (Riaño et al., 2003) and used to estimate crown volume and bulk density (leaves/branches per unit volume). However, one of the most important uses of lidar has been in improved estimates of forest aboveground biomass (Figure 4). Poor knowledge of tropical forest biomass is one of the most important sources of uncertainty in estimating carbon emissions from deforestation (Houghton et al., 2001). Strong correlations between tree height and biomass make

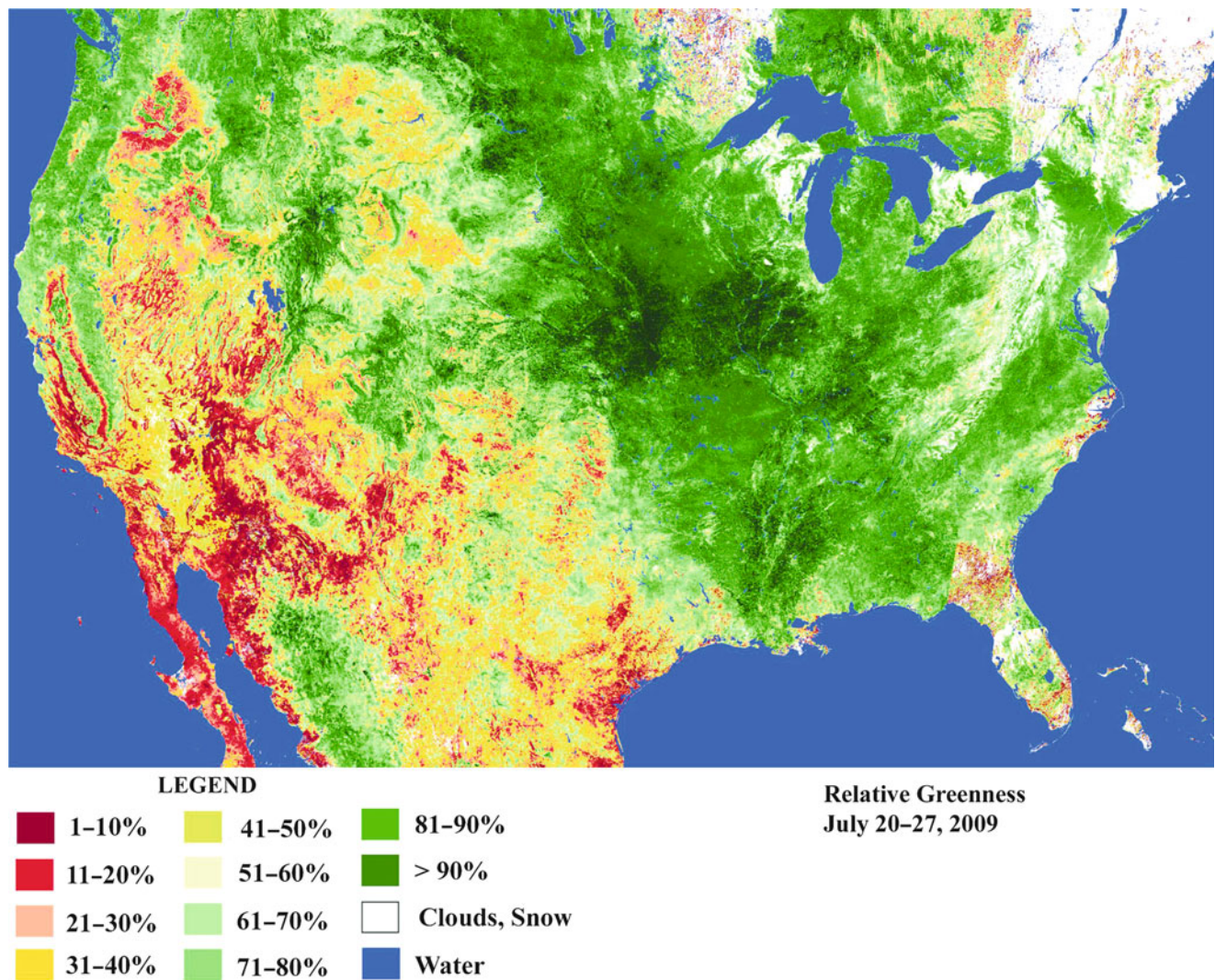


Forestry, Figure 4 Plot of biomass (y) against mean lidar height (x) from four different forests: Wind River (mixed western hemlock and Douglas fir), La Selva (tropical rain forest), Yellowstone (Lodgepole pine), and SERC (mixed broadleaf deciduous). SERC included courtesy of Keely Roth of UC Santa Barbara.



Forestry, Figure 5 Plot showing spectra of important canopy fuels including green leaves (GV) and litter, branches and stems (NPV). Plant litter, branches and stems are readily distinguished from soils based on their spectral shape. As LFM declines, NIR canopy reflectance decreases, chlorophyll absorption becomes muted, and SWIR reflectance increases. Arrows point to important wavelengths used to estimate changes in canopy moisture.

it possible to estimate biomass from lidar height returns. Lidar intensity can also be used to provide species compositional information, especially when given a combination of leaf-on and leaf-off data (Kim et al., 2009). While



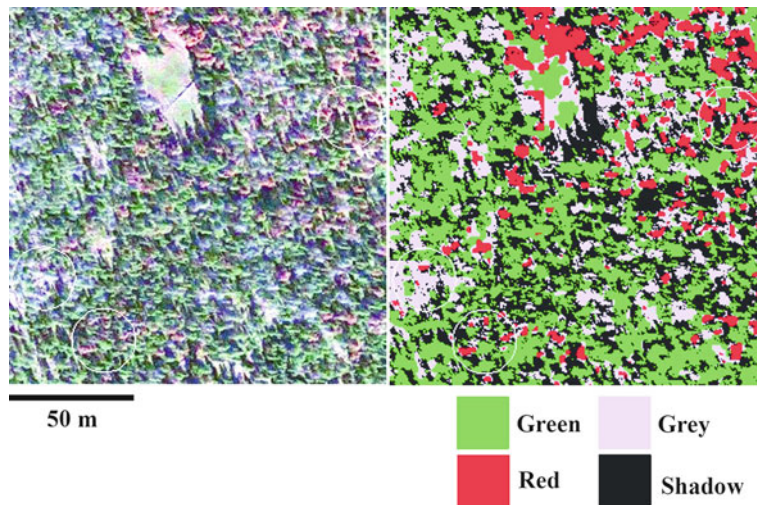
Forestry, Figure 6 Relative greenness for July 20–27, 2009. Areas of below average greenness are shown in various shades of red, and areas with above average greenness are shown in various shades of green (Data from <ftp://ftp2.fs.fed.us/pub/ndvi/>, retrieved from <http://wfas.fire.org/content/view/30/47/>).

a majority of forestry applications have relied on discrete-return lidar, similar high-quality height and biomass estimates have been derived from waveform lidar systems, such as the airborne Laser Vegetation Imaging Sensor (LVIS) (Drake et al., 2002) and the Geoscience Laser Altimeter System (GLAS) onboard ICESAT (Lefsky et al., 2005). Waveform lidar is particularly important because it can be deployed in space and used to estimate biomass from forests that cannot be imaged using airborne systems.

Wildfire fuels and forest fires

Another major remote sensing application in forestry is fire, where remote sensing provides information on fuels, fire danger, fire occurrence, and fire impacts. Improved

knowledge of changing fuel properties and the global incidence of fire is widely considered one of the greatest weaknesses in current global climate models and the ability to quantify fluxes of carbon between the biosphere and atmosphere (Running, 2008). Fires require fuels to burn and are limited by the presence of water; thus, important fuel properties include the mass of fuel, their arrangement in space (depth and packing), their size distribution and geometry, and the amount of water in living and dead tissue (Pyne et al., 1996). Forests can be divided into two important fuel classes, crown and surface fuels. The most extreme fires burn through crown fuels. A common practice is to assign fuel properties using a fuel model, which is typically mapped using image classification. Two of the most common systems used in the United States were proposed by Anderson (1982) and



Forestry, Figure 7 Showing a high-resolution true color pan-sharpened image from Geo-Eye-1 (*left*) and a map of *green*, *red*, and *gray* attack phases of mountain pine beetle generated using a maximum likelihood classifier (*right*) (Adapted from Dennison et al., (2010)).

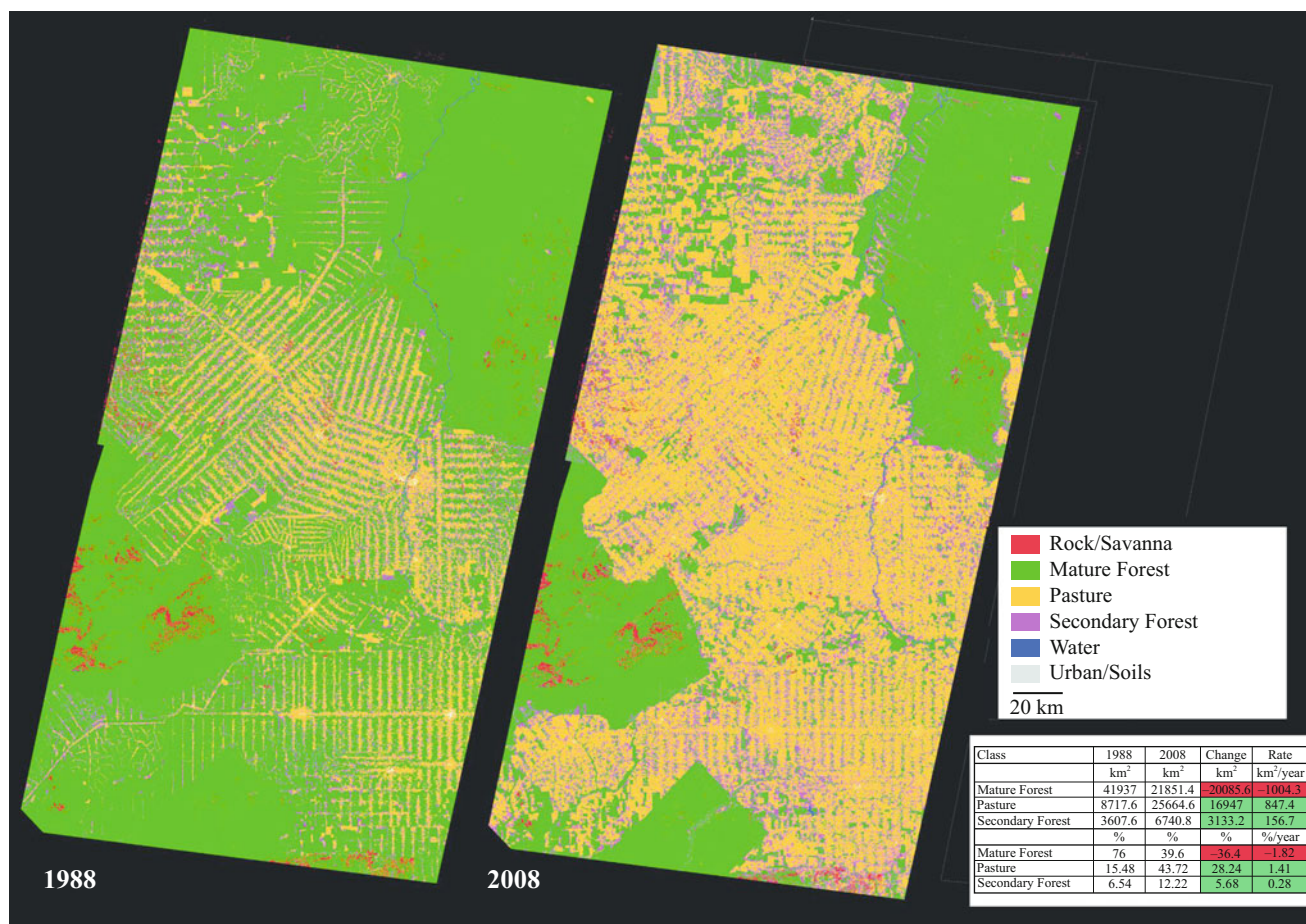
Scott and Burgan (2005), consisting of 13 and 40 models, respectively. These models can be mapped using a variety of passive optical remote sensing systems but have most often been mapped using Landsat Thematic Mapper (TM) data. Because passive optical systems only image canopy tops, they cannot measure surface fuels obscured by the canopy or other important structural variables. However, these can be estimated statistically (Keane et al., 2001). This is the approach currently used by the Landfire program (<http://www.landfire.gov/>) to estimate understory fuels, canopy bulk density, tree height, and crown base height from Landsat TM and biophysical gradients (i.e., aspect, elevation), trained using field data and fire history.

In shrubland systems, many important fuel properties can be estimated directly from remote sensing because the canopy is the primary fuel (Roberts et al., 2003a; Figure 5), including live fuel moisture (LFM: Dennison et al., 2003) and fuel condition (live to dead fuel ratio, Roberts et al., 2003a). A number of sensors have been used to estimate LFM, including Landsat TM (Chuvieco et al., 2002), AVIRIS (Dennison et al., 2003), and the Moderate Resolution Imaging Spectrometer (MODIS: Peterson et al., 2008). MODIS is particularly important in that these data provide information on how LFM changes seasonally.

Prefire fuels are static, typically assigned to a fuel model where the only dynamic element is dead fuel moisture estimated from meteorology. Fire danger, however, is dynamic, varying with changing vegetation and atmospheric conditions. Most fire danger indices, such as the Keetch-Byrum Index (Keetch and Byrum, 1968), rely exclusively on meteorology. However, dynamic changes in fuels can be determined using remote sensing. One example is the use of MODIS to estimate seasonal changes

in LFM (Peterson et al., 2008). Another is relative greenness (RG: Burgan et al., 1998), in which the greenness of a pixel at a specific time is compared to its historical range derived from a long time series. RG is routinely calculated by the US Forest Service using the Normalized Difference Vegetation Index (NDVI) applied to Advanced Very High Resolution Radiometer (AVHRR) as part of its Wildfire Assessment program (WFAS: <http://wfas.fire.org/content/view/30/47/>; Figure 6). RG is also used as an input into the Fire Potential Index (FPI), a danger index calculated from RG and fuel model derived 10 h fuels and dead fuel moisture of extinction (Burgan et al., 1998).

An important element of fire fighting and forest-carbon accounting is the ability to map the presence of fire. Active fire mapping includes fire detection and estimation of important fire properties, such as fire temperature, area, or fire radiative power (FRP). Satellite-based fire detection relies primarily on the Mid-IR, in which fires emit substantially more electromagnetic radiation than cooler background pixels or reflected solar radiation. Example Mid-IR fire detections include daily 1 km observations using AVHRR (3.55–3.93 μm : Setzer and Malingreau, 1996), sub-hourly 4 km observations using Geostationary Operational Environmental Satellite (GOES, 3.9 μm band; Menzel and Prins, 1996), and two or more daily observations from MODIS, an active fire product similar to AVHRR that takes advantage of additional wavelengths and spatial information to reduce false detections (Giglio et al., 2003). Another important MODIS fire product is FRP, an index correlated to biomass consumption (Wooster et al., 2005), calculated using a modified form of Stefan-Boltzmann's equation applied to the difference between background radiance and fire-elevated radiance in the MODIS 4 μm band (Kaufman et al., 1998). Fire temperature and area can be estimated based on the



Forestry, Figure 8 Maps showing changes in mature forest, pasture, and secondary forest in a portion of Rondonia, Brazil, between 1988 and 2008. Landsat TM data were classified using an approach described by Roberts et al., 2002, modified to reduce errors in secondary forest due to illumination and confusion between mature forest and old secondary forest. The 2008 image is included, courtesy of Michael Toomey of UC Santa Barbara. In these two Landsat scenes, mature forest declined from 76 % (41,940 km²) to 39.6 % (21,851 km²), while pasture increased from 15.5 % (8,720 km²) to 43.7 % (25,665 km²) of the area shown. Secondary forests nearly doubled from 6.5 % (3,600 km²) to 12.2 % (6,740 km²) over this same period.

spectral shape of fire-emitted radiance measured by several sensors, including AVIRIS (Dennison et al., 2006).

Immediately following a wildfire, critical questions include how much area was burned and the severity of damage. Measures of burned area and fire severity are critical in that exposed soil following wildfire is prone to erosion and mass movement, while the potential of recovery may vary depending on the intensity and frequency of fire (Barro and Conard, 1991; Zedler et al., 1983). Common fire severity measures are typically calculated using sensors such as Landsat and include the Difference Normalized Burn Ratio (DNBR) as an example (Key and Benson, 2006; Van Wagendonk and Lutz, 2007). DNBR responds to spectral changes due to an increase in exposed soil, surface ash, and dead plant material. Global estimates of burned area are also critical in that

they capture the global incidence of fire disturbance and provide a means for estimating carbon emissions from burned areas. A number of global burned area products exist including one produced using MODIS (Roy et al., 2005).

Forest health

Spectroscopic changes in leaf reflectance (i.e., pigment damage, water loss), and forest structural changes, such as leaf shedding or defoliation, offer the potential of using remote sensing to map the presence of forest pathogens (Pu et al., 2008). Remote sensing has a long history of use for mapping forest health, initially through the use of aerial photography, but more recently using airborne and spaceborne passive remote sensing systems (Wulder

et al., 2006). Examples include the use of multispectral remote sensing to map defoliation due to various species of budworms (Radeloff et al., 1999), adelgids (Franklin et al., 1995), bark beetles (Wulder et al., 2006; Dennison et al., 2010: Figure 7), and gypsy moth (Townsend et al., 2004). Multispectral and hyperspectral remote sensing have been used to map the distribution of forest pathogens such as *Phytophthora*, which causes diseases such as sudden oak death syndrome (Pu et al., 2008) and Swiss needle cast.

Land-cover mapping, forest degradation, and forest conversion (deforestation)

One of the most important roles of remote sensing for forestry has been in tracking changes in the world's forests, either through forest removal (deforestation), degradation, regrowth, afforestation (tree planting), or natural disturbance (i.e., Hurricane Katrina). Forest areas, consisting of at least 40 % canopy cover by woody plants taller than 5 m, increased an average of three million ha/year between 1990 and 2000 in temperate regions, but declined by an average of 12 million ha/year between 1980 and 2000 in tropical regions (Millennium Ecosystem Assessment, 2005). Much of this information has come from passive remote sensing systems, such as the Landsat TM, MODIS, AVHRR, and SPOT-4. Loss of forest cover underestimates actual human impacts because it does not take into account forest degradation, in which forest integrity is reduced by fire, selective logging, or fragmentation. In the Amazon, annual increases in degraded forests are estimated to exceed deforestation (Nepstad et al., 1999).

One of the most intensively studied areas of the humid tropics is the Amazon Basin, which contains the largest remaining tracts of undisturbed tropical rainforest globally and in which a majority of the deforestation has occurred during the era of routine satellite observations (Roberts et al., 2003b). Tropical rain forests are particularly important because they house some of the world's highest biodiversity and store more carbon than any other terrestrial ecosystem (Dixon et al., 1994). Within the Amazon Basin, some of the highest deforestation rates have occurred in the state of Rondonia, which began to undergo rapid conversion as early as the mid-1970s and continues to experience rapid forest loss today (Figure 8).

Summary

Remote sensing has numerous applications in forestry. Older, historical applications relied primarily on interpretation of aerial photography for forest inventories. More recently, newer airborne and spaceborne active and passive sensors have greatly expanded the utility of remote sensing in forestry. Examples include lidar, which can provide forest structural parameters such as tree height, crown volume, and stand biomass at accuracies that are equal to, or exceed, ground-based measures. Regional to global estimates of changes in forest cover have also been made possible by remote sensing,

providing critical measures of changes in carbon stocks and fluxes needed to better understand human environmental impacts and better manage resources. Several operational remote sensing products contribute to improved assessment of fire danger, fire management, fire severity, and postfire recovery.

Bibliography

- Anderson, H. E., 1982. *Aids to Determining Fuel Models for Estimating Fire Behavior*. General Technical Report INT-122. Ogden: United States Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, 26 p.
- Asner, G. P., Martin, R. E., Ford, A. J., Metcalfe, D. J., and Liddel, M. J., 2009. Leaf chemical and spectral diversity in Australian tropical forests. *Ecological Applications*, **19**(1), 236–253.
- Avery, T. E., and Burkhart, H. E., 1983. *Forest Measurements*, 3rd edn. New York: McGraw-Hill, p. 331.
- Barro, S. C., and Conard, S. G., 1991. Fire effects on California chaparral systems: an overview. *Environment International*, **17**, 135–149.
- Burgan, R. E., Klaver, R. W., and Klaver, J. M., 1998. Fuel models and fire potential from satellite and surface observations. *International Journal of Wildland Fire*, **8**(3), 159–170.
- Chuvieco, E., Riano, D., Aguado, I., and Cocero, D., 2002. Estimation of fuel moisture content from multitemporal analysis of Landsat Thematic Mapper reflectance data: applications in fire danger assessment. *International Journal of Remote Sensing*, **23**, 2145–2162.
- Clark, M. L., Clark, D. B., and Roberts, D. A., 2004. Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sensing of Environment*, **91**, 68–89.
- Davis, F. W., and Roberts, D. A., 1999. Stand structure in terrestrial ecosystems. In Sala, O. E., et al. (eds.), *Methods in Ecosystem Science*. New York: Springer, pp. 7–30.
- Dennison, P. E., Roberts, D. A., Thorgusen, S. R., Regelbrugge, J. C., Weise, D., and Lee, C., 2003. Modeling seasonal changes in live fuel moisture and equivalent water thickness using a cumulative water balance index. *Remote Sensing of Environment*, **88**(4), 442–452.
- Dennison, P. E., Charoensiri, K., Roberts, D. A., Peterson, S. H., and Green, R. O., 2006. Wildfire temperature and land cover modeling using hyperspectral data. *Remote Sensing of Environment*, **100**, 212–222.
- Dennison, P. E., Brunelle, A. R., and Carter, V. A., 2010. Assessing red and gray attack canopy mortality caused by mountain pine beetle using GeoEye-1 high spatial resolution satellite data. *Remote Sensing of Environment*, **114**, 2431–2435.
- Dixon, R. K., Brown, S., Houghton, R. A., Solomon, A. M., Trexler, M. C., and Wisniewski, J., 1994. Carbon pools and flux of global forest eco-systems. *Science*, **263**(5144), 185–190.
- Drake, J. B., Dubayah, R. O., Knox, R. G., Clark, D. B., and Blair, J. B., 2002. Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest. *Remote Sensing of Environment*, **81**, 378–392.
- Franklin, S. E., Waring, R. H., McCreight, R. W., Cohen, W. B., and Fiorella, M., 1995. Aerial and satellite sensor detection and classification of western spruce budworm defoliation in a subalpine forest. *Canadian Journal of Remote Sensing*, **21**, 299–308.
- Gates, D. M., Keegan, H. J., Schleter, J. C., and Weidner, V. R., 1965. Spectral properties of plants. *Applied Optics*, **4**(1), 11–20.
- Giglio, L., Descloitres, J., Justice, C. O., and Kaufman, Y., 2003. An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment*, **87**, 273–282.

- Heller, R. C., and Ulliman, J. J., 1983. Forest resource assessment, Chapter 34. In *Manual of Remote Sensing*, 2nd edn. Falls Church: American Society of Photogrammetry, Vol. 2, pp. 2229–2324.
- Houghton, R. A., Lawrence, K. T., Hackler, J. L., and Brown, S., 2001. The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates. *Global Change Biology*, **7**, 731–746.
- Kaufman, Y. J., Justice, C. O., Flynn, L. P., Kendall, J. D., Prins, E. M., Giglio, L., et al., 1998. Potential global fire monitoring from EOS-MODIS. *Journal of Geophysical Research*, **103**, 32215–32238.
- Keane, R. E., Burgan, R., and van Wagtenonk, J., 2001. Mapping wildland fuels for fire management across multiple scales: integrating remote sensing, GIS, and biophysical modeling. *International Journal of Wildland Fire*, **10**(3–4), 301–319.
- Keetch, J. J., and Byram, G. M., 1968. *A Drought Index for Forest Fire control*. Asheville: Southeastern Forest Experiment Station. U.S. Department of Agriculture, Forest Service Research Paper SE-38.
- Key, C. H., and Benson N., 2006. *Landscape assessment: sampling and analysis methods*. Report RMRS-GTR-164-CD, USDA. Washington, DC: Forest Service.
- Kim, S., McGaughey, R., Andersen, H.-E., and Schreuder, G., 2009. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. *Remote Sensing of Environment*, **113**(8), 1575–1586.
- Lefsky, M. A., Cohen, W. B., Parker, G. G., and Harding, D. J., 2002. LIDAR remote sensing for ecosystem studies. *Bioscience*, **52**, 19–30.
- Lefsky, M. A., Harding, D. J., Keller, M., Cohen, W. B., Carabajal, C. C., Del Bom Espirito-Santo, F., Hunter, M. O., and de Oliveira, R., Jr., 2005. Estimates of forest canopy height and aboveground biomass using ICESat. *Geophysical Research Letters*, **32**, L22S02, doi:10.1029/2005GL023971.
- Menzel, W. P., and Prins, E. M., 1996. Monitoring biomass burning with the new generation of geostationary satellites. In Levine, J. S. (ed.), *Biomass Burning and Global Change*. Cambridge: MIT Press, Vol. 1, Chap. 6, pp. 56–72.
- Millennium Ecosystem Assessment, 2005. Ecosystems and human well-being synthesis. <http://www.millenniumassessment.org/documents/document.356.aspx.pdf>, pp. 155.
- Nepstad, D. C., Verissimo, J. A., Alencar, A., Nobre, C., Lima, E., Lefebvre, P., Schlesinger, P., Potter, C., Moutinho, P., Mendoza, E., Cochrane, M., and Brooks, V., 1999. Large-scale impoverishment of Amazonian forests by logging and fire. *Nature*, **398**, 505–508.
- Peterson, S. H., Roberts, D. A., and Dennison, P. E., 2008. Mapping live fuel moisture with MODIS data: a multiple regression approach. *Remote Sensing of Environment*, **112**(12), 4272–4284.
- Pu, R., Kelly, M., Anderson, G. L., and Gong, P., 2008. Using CASI hyperspectral imagery to detect mortality and vegetation stress associated with a new hardwood forest disease. *Photogrammetric Engineering and Remote Sensing*, **74**(1), 65–75.
- Pyne, S. J., Andrews, P. L., and Laven, R. D., 1996. *Introduction to Wildland Fire*. New York: Wiley.
- Radeloff, V. C., Mladenoff, D. J., and Boyce, M. S., 1999. Detecting jack pine budworm defoliation using spectral mixture analysis: separating effects from determinants. *Remote Sensing of Environment*, **69**, 156–169.
- Riaño, D., Meier, E., Allgöwer, B., Chuvieco, E., and Ustin, S., 2003. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment*, **86**, 177–186.
- Roberts, D. A., Numata, I., Holmes, K. W., Batista, G., Krug, T., Monteiro, A., Powell, B., and Chadwick, O., 2002. Large area mapping of land-cover change in Rondônia using multitemporal spectral mixture analysis and decision tree classifiers. *Journal of Geophysical Research-Atmospheres*, **107**(D20), 8073. LBA 40–1 to 40–18.
- Roberts, D. A., Dennison, P. E., Gardner, M., Hetzel, Y., Ustin, S. L., and Lee, C., 2003a. Evaluation of the potential of hyperion for fire danger assessment by comparison to the airborne visible/infrared imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6 Part 1), 1297–1310.
- Roberts, D. A., Keller, M., and Soares, J. V., 2003b. Studies of land-cover, land-use, and biophysical properties of vegetation in the large scale atmosphere experiment in Amazonia (LBA). *Remote Sensing of Environment*, **87**(4), 377–388.
- Roberts, D. A., Ustin, S. L., Ogunjemiyo, S., Greenberg, J., Dobrowski, S. Z., Chen, J., and Hinckley, T. M., 2004. Spectral and structural measures of Northwest forest vegetation at leaf to landscape scales. *Ecosystems*, **7**, 545–562.
- Roy, D. P., Jin, Y., Lewis, P. E., and Justice, C. O., 2005. Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sensing of Environment*, **97**(2), 137–162.
- Running, S. W., 2008. Ecosystem disturbance, carbon, and climate. *Science*, **321**, 652–653.
- Salisbury, J. W., and Milton, N. M., 1988. Thermal infrared (2.5 to 13.5- μ m) directional hemispherical reflectance of leaves. *Photogrammetric Engineering and Remote Sensing*, **54**, 1301–1304.
- Scott, J. H., and Burgan, R. E., 2005. *Standard Fire Behavior Fuel Models: A Comprehensive Set for Use With Rothermel's Surface Fire Spread Model*. Fort Collins: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. General Technical Report RMRS-GTR-153, 72 p.
- Setzer, A. W., and Malingreau, J. P., 1996. AVHRR monitoring of vegetation fires in the tropics: toward the development of a global product, Chapter 3. In Levine, J. S. (ed.), *Biomass Burning and Global Change*. Cambridge: MIT Press, Vol. 1, pp. 25–39.
- Townsend, P. A., Eshleman, K. N., and Welcker, C., 2004. Remote sensing of gypsy moth defoliation to assess variations in stream nitrogen concentrations. *Ecological Applications*, **14**, 504–516.
- van Wagtenonk, J. W., and Lutz, J. A., 2007. Fire regime attributes of wildland fires in Yosemite National Park, USA. *Fire Ecology*, **3**(2), 34–52.
- Wooster, M. J., Roberts, G., Perry, G. L. W., and Kaufman, Y. J., 2005. Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release. *Journal of Geophysical Research*, **110**, D24311, doi:10.1029/2005JD006318.
- Wulder, M., 1998. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. *Progress in Physical Geography*, **22**, 449–476.
- Wulder, M. A., Dymond, C. C., White, J. C., Leckie, D. G., and Carroll, A. L., 2006. Surveying mountain pine beetle damage of forests: a review of remote sensing opportunities. *Forest Ecology and Management*, **221**(1–3), 27–41.
- Zedler, P. H., Gautier, C. R., and McMaster, G. S., 1983. Vegetation change in response to extreme events: the effect of a short interval between fires in California chaparral and coastal scrub. *Ecology*, **64**, 809–818.

Cross-references

Data Processing, SAR Sensors
 Lidar Systems
 Microwave Radiometers
 Optical/Infrared, Radiative Transfer
 Radars
 Radiation, Electromagnetic
 Reflected Solar Radiation Sensors, Multiangle Imaging

G

GAMMA AND X-RADIATION

Enrico Costa and Fabio Muleri
Istituto di Astrofisica e Planetologia Spaziali, INAF,
Rome, Italy

Synonyms

Gamma radiation; Gamma rays; High-energy photons;
High-energy radiation; X-radiation; X-rays; γ -rays

Definition

X-radiation. Electromagnetic radiation with wavelength λ shorter than ultraviolet light but longer than γ -rays, $0.01 \text{ nm} \leq \lambda \leq 10 \text{ nm}$.

Gamma radiation. Electromagnetic radiation with energy higher (or, equivalently, shorter wavelength/higher frequency) than X-radiation.

Introduction

X-rays and gamma rays are the shorter wavelength or, equivalently, the higher-energy side of the electromagnetic spectrum. The separation of these two bands is quite conventional: X-rays are considered to have an energy from that of ultraviolet light to the binding energy of inner electronic shells of atoms, whereas above this value there are gamma rays which, broadly speaking, are produced by nuclear activity. For the sake of simplicity, we name as X-radiation those electromagnetic waves with wavelength λ between 10 and 0.01 nm, while gamma rays are assumed to have a value of λ below 0.01 nm. A common way of describing gamma and X-radiation is by the energy expressed in units of electron volt (eV), where 1 eV is the energy acquired by an electron which passes through a difference of potential of 1 V. With this definition, X-rays have energies between about 100 eV and 100 keV (1 keV = 1,000 eV). The relation between the

wavelength in nm and the energy of photons expressed in eV is given by:

$$E = \frac{1239.84}{\lambda_{nm}} \text{ eV}. \quad (1)$$

X-rays were discovered by Wilhelm Roentgen in 1895 as a new kind of emission obtained by the interaction of accelerated cathode rays, that is, electrons, with a target (Shamos, 1987). Their characteristics were a relatively long range and the propagation in straight lines, even in the presence of a magnetic field, which suggested to rule out the hypothesis that X-rays were charged particles. However, it was not earlier than 1912 that the real nature of X-rays as short wavelength electromagnetic radiation came out. Max von Laue showed that X-rays are diffracted by crystal lattices thus proving that X-rays are actually electromagnetic radiation with wavelength comparable to lattice spacing. Gamma rays were discovered by Paul Villard in 1900, who recognized the presence, beyond the α -rays (helium nuclei) and β -rays (electrons), of a component unaffected by magnetic field and scarcely absorbed in the emission of radioactive materials (Gerward, 1999). Also in this case the definite proof that gamma rays are photons with short wavelength came from their diffraction on crystal lattice.

Interaction of X-rays and gamma rays with matter

The short wavelength makes the behavior of gamma and X-radiation much similar to that of discrete particles. Notwithstanding, there is a fundamental difference in the interaction with matter, that is, photons are destroyed when they interact whereas charged particles lose energy continuously.

There are three main processes through which gamma and X photons interact with matter and the probability of occurring of each depends on the energy of the radiation.

In the low-energy part of the X-ray band (soft X-rays), the prevailing interaction is with electrons in inner shells of atoms via the photoelectric effect (or photoabsorption). An electron (named photoelectron) is extracted from the atom with a kinetic energy equal to the difference between the energy of the photon and the binding energy of the electron. Only interactions with electrons bound with an energy lower than that of the photon are possible, and therefore, inner electronic shells are being involved only for increasing photon energy. The contribution of the most internal shell is always predominant with respect to the outer ones. The probability of interaction with a certain shell, that is, the cross section of the process σ , is a steep decreasing function of energy, $\sigma \sim E^{-3.5}$. This dependency is characterized by an abrupt increase of σ as soon as the energy is sufficient to extract photoelectrons from a more inner shell (*absorption edge*). The probability of photoelectric absorption has a strong dependence on the atomic number Z ($\sigma \sim Z^5$) and thus, heavy elements absorb photons with a much more efficiency. The ejection of the photoelectron leaves the ionized atom in an excited status with a vacancy in an inner shell. It can decay via two major processes: the ejection of a fluorescence photon or the ejection of a self-ionization electron (named Auger electron). In both cases, the total energy released is just below the binding energy of the photoelectron. The detection of fluorescence emission is a powerful tool for determining the composition of materials because each element produces radiation with a characteristic pattern, that is, only at those energies which correspond to transitions between its own electronic shells. Therefore, by measuring the energy and the intensity of fluorescence photons, it is possible to identify the atomic elements and their abundance without destroying the sample under test.

At energies higher than the binding energy of most inner electrons (belonging to the K-shell), the probability of photoabsorption decreases quickly. When the energy is tens of times the K-shell energy, namely, between a few tens to a few hundreds of kilo electron volts depending on the atomic number of the material, the prevailing interaction of photons becomes the Compton effect. In this case, the photon is scattered off an electron and the former transfers a part of its energy to the latter. Even if electrons in materials are bound in atoms, for what concern the process of interaction they can be regarded as to be free, because the binding energy is in any case negligible with respect to the energy of the photon. By the conservation of the energy and of the momentum, the energy of the scattered photon E' can be calculated as a function of the scattering angle θ from the energy of the incident photon E :

$$E' = \frac{E}{1 + \frac{E}{mc^2}(1 - \cos \theta)}. \quad (2)$$

The energy acquired by the electron is the difference between E and E' . The probability of scattering and the

angular distribution of photons are described by the Klein-Nishina formula:

$$\frac{d\sigma}{d\Omega} = \frac{1}{4} \left(\frac{e}{mc^2} \right)^2 \frac{E'^2}{E^2} \left(\frac{E}{E'} + \frac{E'}{E} - \sin^2 \theta \right). \quad (3)$$

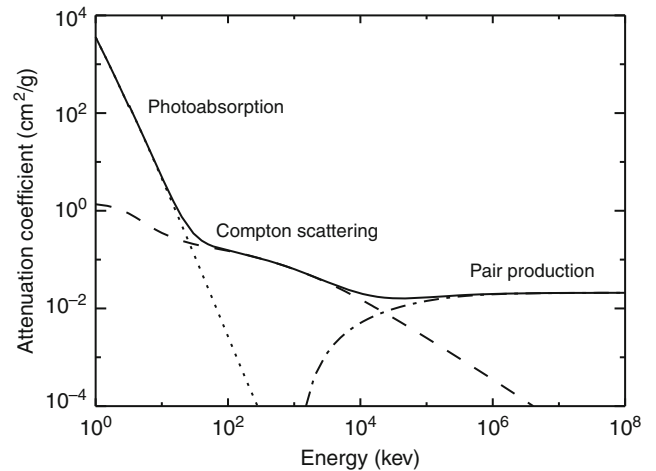
Another process of interaction between photons and matter is possible, that is, the pair production. If the photon has an energy higher than two times the rest mass of electrons (>1.022 MeV, $1 \text{ MeV} = 1,000 \text{ keV}$), the photon can split into an electron-positron pair. A third body, usually the field of a nucleus but also that of an electron, must be present to allow for the conservation of both energy and momentum, and it can absorb a part of the initial photon energy. Above some tens of MeV, pair production is the most probable interaction and the cross section becomes basically independent from energy.

Interaction of X-rays and gamma rays with Earth's atmosphere

Earth's atmosphere strongly absorbs high-energy radiation, especially in the soft X-ray range. The fraction of photons which pass through it without interacting, that is, the transparency T , as a function of energy is calculated by using the attenuation coefficient μ , whose value can be found in the literature (see, e.g., the "XCOM: Photon Cross Sections Database," Berger et al.). If the depth of the medium is x and ρ is its density, we have:

$$T = \exp(-\mu\rho x). \quad (4)$$

In Figure 1, we report the attenuation coefficient for the Earth atmosphere as a function of energy. The absorption



Gamma and X-Radiation, Figure 1 Attenuation coefficient as a function of energy for Earth's atmosphere. The contributions of photoabsorption (dotted line), of Compton scattering (dashed line), and of pair production (dash-dot line) are distinguished. It is assumed a composition of N₂ (78.1 %), O₂ (20.9 %), and Ar (0.01) (NASA's Earth Fact Sheet). The value of the attenuation coefficient is retrieved from XCOM database (Berger et al.).

is dominated by photoelectric effect, Compton scattering, or pair production at different energies, but its value remains significant all through the spectrum. For example, the attenuation length, defined as the distance at which the transmission is reduced to $1/e$, is 0.23 cm at 1 keV, and the transparency of 1 cm is about 1 % (the density of the atmosphere is assumed to be $1.2 \times 10^{-3} \text{ g/cm}^3$). At higher energies, the absorption is less effective but even at the minimum of the attenuation coefficient around 40 MeV, the attenuation length is only a few hundreds of meters. Moreover, in the range of maximum transparency, the prevailing interaction is the Compton scattering. The radiation interacting with atmosphere is scattered forward with typical angles around 30° and with an energy significantly lower than that of the original photon. Therefore, even if a significant number of photons can survive a thickness of a few kilometers, any spectral or imaging information is lost. As a consequence remote sensing is not an effective tool in these energy ranges for the study of the surface of our planet. This is true for the Earth but not necessarily for other planets: Any celestial body, satellite, or minor body, without an atmosphere, is a potential target for remote sensing in X-/gamma rays – actually a few such studies of remote sensing have been performed and are planned for future space missions.

Detectors for X-/gamma rays

Modern detectors allow to measure X- and gamma radiation with a fine sensitivity (Leo, 1994; Lutz, 2007; Knoll, 2010). Spectroscopy, at the level of 1 eV, and imaging, with a spatial resolution up to 1 μm , are the more developed fields, particularly in the soft X-ray band, but also polarimetry is progressing. X- and gamma rays can be detected only after they have interacted with matter via one of the processes described above, which destroys the initial photon and results in the transfer of its energy to one or more electrons. These electrons propagate in the matter while being scattered by interactions with atomic nuclei and slowed down by interactions with atomic electrons. The energy lost along the path ionizes atoms, producing other electrons, and excites electrons of external shells, which shortly afterward decay and emit UV photons. Electrons extracted by ionization propagate at short distance because they have a very low energy, and hence, the result of the interaction of the photon is a small cloud of ion-electron pairs close to the interaction point and a small flash of UV photons.

We can classify detectors in the following three groups:

Detectors that collect the charges produced by ionizing electrons

One way to detect high-energy photons is to measure the charge deposited in the sensitive volume of the instrument by the electrons produced in consequence of the initial interaction. These charges are produced near the point where the photon interacted and are proportional, in number, to the energy of the photon. Moreover, they

are produced along the path of initial electrons, which contains information on the polarization of the photon. For example, the average energy necessary to the electrons to produce other electron/ion pairs in a gas by ionization is 20–25 eV, and therefore, the energy of a photon of 10 keV is converted in around 400 electrons. An electric field is applied in the sensitive volume of the instrument to drift the electron/ion pairs toward an anode/cathode which collects them. To overcome the noise of read-out electronics, often the pairs are multiplied before collection (of a factor spanning from 10^3 to 10^6 and above) in a region with an intense electric field, which accelerates the electrons so that they can ionize other atoms.

The *proportional counter* was the first device developed for detecting high-energy radiation, capable to measure the energy of individual photons, and it can illustrate very well the functioning of detectors that collect the charges produced by ionizing electrons. The sensitive volume is a chamber filled with a gas, where photons are absorbed and electrons are drifted by an electric field to an anode wire (15–50 μm thick) at high voltage. The field close to the wire is very high and accelerates the electrons that collide with atoms producing more electrons. The original cloud of electron is amplified of a factor 10^3 – 10^5 and is collected by the anode wire, while the ions are collected by other wires or by the box of the detector acting as a cathode. This is the basic function of proportional counters. With the use of many wires, they can act as imaging devices with a resolution down to 0.1 mm. The energy resolution is of the order of 20 % at 6 keV and scales with the square root of the energy.

Semiconductor detectors are another type of instrument which measures the charge produced as the result of the photon interaction. Instead of gas, the sensitive volume is a junction of a crystalline semiconductor (mostly silicon and germanium) inversely polarized. When an ionization cloud is produced inside the depletion layer of the junction, the electrons and holes are immediately collected at the electrodes. Since the energy to produce an electron/hole pair in silicon is 3.4 eV, photons of the same energy produce more charges in a semiconductor than in a gas, and this results in an enhancement of the energy resolution. Another advantage is that solid-state instruments are more compact with respect to gas detectors of comparable efficiency.

Semiconductor detectors are usually used without charge multiplication and hence need very low-noise electronics. This requires a low leakage current, a low capacity of the detector, and an operation at low temperature. Since the depleted layer is the actual sensitive part, it is enhanced by building junctions on two sides of an intrinsic, high-resistivity crystal. The capacity is reduced by reducing the surface of the read-out anode, which means detectors with a smaller surface or instruments which exploit a drift field to collect the charges produced in a large volume on a smaller anode (*silicon drift detector*; see below). Imaging detectors can be built by dividing a detector into regions each one with a dedicated electronics to read out the signal. This can be

done in one dimension with the so-called microstrip detectors or in two dimensions with the so-called pixel detectors. By increasing the number of pixels, the encumbrance, the power consumption, and the difficulty to bond each pixel with its electronic chain with 100 % efficiency has oriented the technology toward an X-ray adaptation of the charge-coupled device (CCD). In this case, the charge produced by each absorbed photon is collected and temporarily stored in the pixel closer to the interaction point, but the read-out of the pixels occurs sequentially with a single low-noise electronics. The potentials of the pixels are modified periodically to repeatedly shift the charge from the pixel to its neighbor and, eventually, to the output where the charge is preamplified and converted from analog to digital. A CCD devoted to X-rays is substantially the same of a CCD for optical application with some minor changes: (1) A window in the entrance prevents the optical photons to reach the sensitive volume, since it would be saturated by the low-energy electrons generated from the absorption of visible/UV photons. (2) The depletion layer is made thicker to increase the efficiency. (3) The dead layers are minimized in order to minimize the absorption of X-rays. While in an optical CCD each photon creates 1 (or 0) electron and the charge collected in the pixel is proportional to the number of impinging photons, a CCD for X-rays is usually discharged very frequently so that the probability to have more than one photon detected in the same pixel is low and the charge measured from the pixel is proportional to the energy of the photon. Therefore, the CCD is an excellent imager (down to resolutions of the order of 10 μm) and an excellent spectroscopic device (down to resolutions of the order of 150 eV). The major limitations are in the surface that cannot exceed a few square centimeters and in the time resolution that cannot be better than a fraction of milliseconds because of the relatively long time required to transfer the charge of the whole matrix of pixels to the output without losses. A recent evolution of CCDs, still in the development stage, is the *advanced pixel sensor* that reads out the charge without transferring it, with a more parallel read-out and a better energy resolution. More in general the CCD suffers the limitations due to the use of silicon as absorbing material, which becomes quite inefficient above some tens of keV, the actual value depending on the thickness. To detect photons of higher energies, detectors have been studied based on materials of higher atomic number such as GaAs and CdTe, but the technologies have not arrived to the level of development of CCDs yet. Other alternative technology is the *silicon drift detectors* where the charge produced is promptly drifted to the output and the coordinate in the drift direction is measured from the drift time, the drift velocity being known. These have a good energy resolution at room temperature but are compatible only with one-dimensional imaging. A further improvement is the development of *avalanche silicon detectors* that, by exploiting the small difference in the mobility of electrons and holes, can perform a proportional multiplication of the charge.

Detectors that measure the light

X-ray photons can also be detected by measuring the light emitted as a consequence of ionization/excitation processes in suitable crystals. These photons are typically in the UV band and are absorbed in few tens of microns, making impossible their direct detection. However, if crystals are doped with elements that create suitable metastable levels, the absorption of the UV photons creates new photons shifted to longer wavelength. The latter are capable to travel across the crystal to an optical window without being absorbed. A photomultiplier with its electronics detects the light, giving an electrical pulse whose height is proportional to the number of photons collected from the cathode and thence to energy released in the interaction.

Scintillators to detect X- and gamma rays are based on inorganic crystals. The most common have been for years NaI (thallium activated), the CsI (thallium or sodium activated), because of good efficiency and high light yield. The $\text{Bi}_4\text{Ge}_3\text{O}_{12}$ (that does not require activation) is most common to detect higher-energy gammas but has a poor light yield. The more recently discovered LaBr_3 (cerium activated) is nowadays the best performing material since it combines a high quantum efficiency with a high light yield that results in a higher-energy resolution. Scintillators are usually employed to detect hard X-rays (>10 keV) and soft gamma rays. They are suitable to achieve large volumes, and smaller crystals can be made position sensitive when coupled with multi-pixel photomultipliers. Larger crystals in the form of bars or disks can be made position sensitive by recording how light is shared by different photomultipliers in different parts of the crystal. Gamma cameras for CAT are a classical application of this method.

Detectors that measure the heat

After the very first interaction, all the energy absorbed from the original photon is eventually converted into heat. If a crystal is small and is kept at very low temperature, where the specific heat is extremely small, also the small amount of energy carried by an X-ray photon is sufficient to produce a measurable increase of the temperature. The microcalorimeters (or X-ray bolometers) are based on this effect. The detector is conceptually composed by an absorber that converts the X-rays, a thermometer in good thermal contact to it and a weak link to a thermal bath to restore the initial condition after the photon absorption. Although the more mature technology is to use semiconductor thermistors to measure the small temperature variation of the absorber, the more promising devices exploit a superconductor kept just below the critical temperature. The heat is sufficient to induce the transition from superconducting to normal state and the change of the current flowing in it is detected by a SQUID (Superconducting Quantum Interference Device). The total energy absorbed is derived from the sequence of the

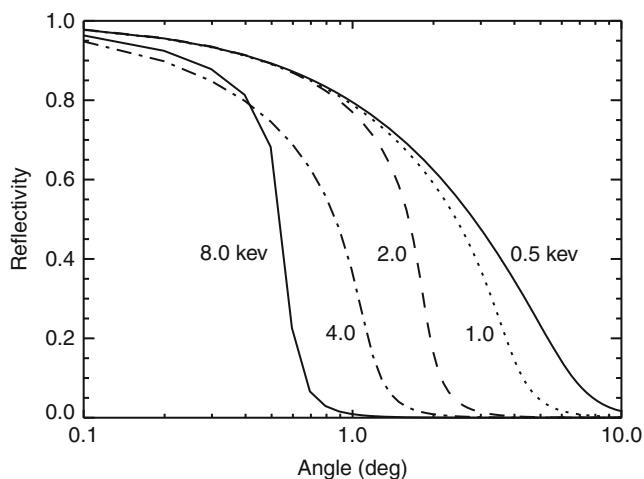
transition as mapped by the SQUID. This is the Transition Edge Sensor, nowadays, the best performing detector in terms of energy resolution (the current record is 1.8 eV at 6 keV). The development of large matrices of TES and of the read-out of individual pixels is the new frontier and the step needed to give practical application to this technology.

Optics and other imaging techniques

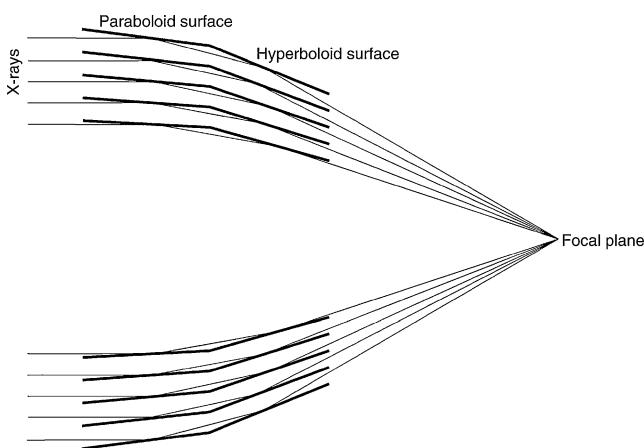
X-ray and soft gamma rays can be focused by means of suitable reflective mirrors like visible light, even if with quite severe limitations. The main difference is that high-energy radiation is reflected only if the grazing angle is below a certain threshold of the order of a few degrees, and this value decreases while increasing energy. This is because the index of refraction at these energies is very close to one. The critical angle above which reflection substantially does not occur is larger for materials of higher atomic number, and indeed it is roughly proportional to the square root of the density of the reflecting material. For this reason a heavy and malleable material as gold is used. As an example, in Figure 2 the reflectivity of a mirror made of gold is reported as a function of the grazing angle and for different energies between 0.5 and 8 keV. Unfortunately, the use of high Z materials causes also a large drop in the reflectivity for energies close to the absorption edges because of the high probability of photoabsorption. This can be filled with a suitable coating on the reflecting material, that is, a material of different atomic number which can reflect X-rays in that particular energy band.

The possibility of reflecting X-rays only for angles below a few degrees implies that the telescope must have a long focal length. This makes the telescopes rather cumbersome. The cost of such a device is also very high, because each reflective surface is to be manufactured with very low roughness, comparable to the wavelength of the radiation to be reflected, precisely aligned and supported by a mechanical structure which provides an adequate stiffness.

X-ray telescopes find their most important use in astronomy since they enable to resolve extended sources and crowded regions, reduce the background, and provide an adequate collecting area. The optics comprises many nested and concentric shells which are formed by two surfaces, the first is a paraboloid and the second is a hyperboloid. The radiation is focused in the focal plane after two reflections on the two surfaces. This configuration, named *Wolter type I*, allows to produce sharp images even off-axis and to reduce the focal length and so the volume (see Figure 3). Inner shells contribute less to the total area because of the smaller radii, but external ones can provide a significant contribution only at low energy since they must reflect photons at larger angles. This actually causes a fast decrease of the effective area with energy, and as a matter of fact, these telescopes can work only below 10–15 keV.



Gamma and X-Radiation, Figure 2 Reflectivity for gold as a function of the grazing angle and for different photon energies. Solid, dotted, dashed, dot-dash, and dash-dot-dot refer to radiation at 0.5, 1.0, 2.0, 4.0, and 8.0 keV, respectively (Data from Henke et al. (1993)).



Gamma and X-Radiation, Figure 3 Sketch of a telescope for X-ray astronomy. Tens of concentric shells are nested and each contributes to the total area, even if external shells can provide a significant contribution only at low energy because of the larger reflection angles required. X-rays are reflected two times on a paraboloid and on a hyperboloid surfaces to provide good imaging capability even off-axis and to reduce the focal length.

The most advanced X-ray telescope is that onboard the Chandra X-ray Observatory, orbiting the Earth since 1999 (see the Chandra X-ray Observatory website). It was designed to achieve the outstanding angular resolution of 0.5 arcsec, and this has required a focal length of 10 m. The telescope comprises of four nested paraboloid/hyperboloid shells which provide a collecting area of 400 cm² at 1 keV. The working energy range is between 0.1 and 10 keV, and only the telescope with the support required to keep aligned the mirrors weighs about 1 t.

A new technology, that is, the multilayer coating, allows to focus X-rays with grazing incidence optics up to hundreds of keV. Instead of a single material with a coating, the shells are made of stacked alternating layers of low and high atomic number materials, deposited via vacuum processes such as evaporation or sputtering. Each layer has a thickness of the order of 1–10 nm which, acting as a crystal with such lattice spacing, diffracts X-rays. Diffraction is a phenomenon which occurs when an electromagnetic wave encounters a spatially repeated structure, such as a grating, with spacing comparable to its wavelength. For X-rays, this occurs with crystal lattices which typically have spacing of the order of a few angstroms (1 Å = 0.1 nm). The planes of the crystal act as reflecting surfaces and the radiation sum, in phase, in those directions which make the difference of optical path between reflections on different planes an integer value of wavelength. Bragg's law relates the energy and the angle at which the diffraction can occur:

$$E = \frac{nhc}{2d \sin \vartheta} \quad (5)$$

where h and c are Planck's constant and the speed of light in vacuum, respectively, d is the crystal lattice spacing, ϑ is diffraction angle, and n an integer which indicates the order of diffraction. Multilayer optics can reflect photons at higher energy, or at larger glancing angles, than "classical" telescopes. By substituting $\vartheta = 1^\circ$ and $d = 1$ nm in Equation 5, it results that the energy diffracted is 35 keV (for $n = 1$). This is a great improvement with respect to single-layer optics, by which photons are reflected at angles larger than 1° only below a few kiloelectron volts. The diffraction from the single layer is nearly monochromatic since, for a fixed value of the spacing of the layer d , the energy is also fixed except for the diffraction order. To diffract continuum spectra, the thickness of the layers is gradually decreased with depth so that the total response is the convolution of nearly monochromatic lines at different energies. Multilayer optics in X-ray astronomy was tested for the first time between 5 and 80 keV by the NASA mission NuSTAR, which was launched in 2012 (Harrison et al., 2013).

Above the hard X energy range, the imaging of radiation cannot be performed with multilayer optics, because it is difficult to make layer with thickness below 1 nm and low roughness. The most promising technique to concentrate soft gamma rays is the use of Laue lenses: photons pass through crystals and are diffracted at grazing angles on the lattice, whose spacing is of the order of a few angstroms (Laue geometry). Laue lenses actually do not focus the image on a focal plane but they just concentrate the radiation. Although the image of a point source on-axis is really a punct, the quality of the image basically depending on the alignment of crystals which compose the lens, it appears like a ring off-axis. Since diffraction on crystal is nearly monochromatic, the efficiency for continuum spectra is increased, at the expense of the

image quality, with the use of mosaic crystals, which are composed of small domains slightly and regularly misaligned acting as independent crystals. Laue lenses are at the moment in development stage, and their use can be expected in the near future.

Imaging of hard X- and gamma rays can be performed also with *coded masks*. These are an evolution of the simple pinhole cameras: The direction of photons is reconstructed by the shadow projected on a sensitive plane. A coded mask instrument is composed of a mask, which is divided in elements, and of a position-sensitive detector. The elements are arranged in a predetermined pattern; some of them are transparent to the radiation to be detected, while the others are opaque. The shift of the mask shadow with respect to center and the distance between the detection plane and the mask enable to reconstruct the direction of impinging photons. The former is measured with a detector with a spatial resolution better than the size of the mask elements. If there is more than one source, the shadows projected by each are summed but the arrangement of the mask elements is chosen so that in any case the position of the sources can be reconstructed without ambiguity by means of a deconvolution process. The intensity of each source is proportional to the "strength" of the corresponding shadow, while the angular resolution is dictated by geometrical parameters, such as the spatial resolution of the detector, the size of the mask elements, and the distance with the detection plane. As a matter of fact, angular resolutions are limited to a few arc minutes because better values would require long focal length and hence quite large volumes.

Coded masks provide a low cost and compact solution to image high-energy radiation up to soft gamma rays, even if the sensitivity of a focusing instrument is always better for an equal collecting area. A coded mask detector measures as a whole the shadow, and hence the noise, of all sources. As a result, the noise which affects the single source is the sum of all contributions, while focusing instruments resolve them spatially. Moreover, the actual geometrical area of the detector in a coded mask device is always larger than its collecting area because about a half of flux must be absorbed by the mask. To reach a large collecting area, an even larger sensitive area is required and this implies a higher background. On the contrary, the flux of sources detected by instruments inserted in focal planes are affected only by the background in the small region where photons are concentrated, that is, the point spread function of the telescope. An application where coded masks are largely used is the imaging of a large fraction of the sky. Grazing incidence telescopes have a field of view smaller than 1 square degree, while a single-coded mask instrument can cover even more than one tenth of the sky, and hence, the entire sky coverage can be performed with a small cluster of detectors.

Another interesting possibility to build large field of view and sensitive instruments is the use of *lobster eye* optics, which is based on the same principle as the

eye of crustaceans (Angel, 1979). The optics is composed of an array of square channels, which are arranged so that their axes are orthogonal to a spherical surface with a radius R . The internal surfaces of channels are made reflective and a true image is formed by those rays which experience an odd number of reflections on both the couples of parallel walls of the channel. Lobster eye lenses are optically equivalent to a concave spherical mirror of radius R and indeed the focal plane is also spherical, with a radius equal to $R/2$. Photons which are reflected an even number of times, also if in just one direction, produce a linear image, parallel to the sides of the channels and passing through the focused image, and, together with photons which are not reflected, are to be regarded as an additional noise. The diagonal of the channel is also the diameter of a point source image on the focal plane, because rays focused by any channel emerge parallel from it with a beam size equal to the diagonal. In principle, lobster eye optics can have a field of view as large as desired up to 4π steradian, that is a complete angular coverage, because there is not any preferred direction, but actually there are technical limitations to build spherical optics and detectors. Also the angular resolution can be very good, and it can be of the order of arc minutes or below. It is basically determined by the ratio between the size of the channel and the focal length, that is, $R/2$.

Lobster eye telescopes are built by exploiting slumped lead glass microchannel plates (MCP, Fraser et al., 1993). Coatings such as iridium or nickel are used to enhance the reflectivity, and the reflection up to hard X-rays has been reported (Price et al., 2002a). An improvement to the basic design is also the packing of a pair of MCPs to form an approximate Wolter type I optics, with each MCP composed of square channels arranged radially (Willingale et al., 1998; Price et al., 2002b). This design improves the performance from many points of view, such as better imaging capabilities and higher effective area, but the field of view is narrower because a preferred direction is introduced by the system.

Recently, lobster eye optics has been also developed with metal ribs coated with gold (Gertsenshteyn et al., 2005). The possibility to polish the walls of the channels allows for the diffraction up to hard X-rays (up to 40 keV) and has opened the way for a number of applications. One of the most interesting is the possibility to inspect objects in a fast and nondestructive way with a portable instrument. This comprises of an X-ray generator, which illuminates the material, and a detector in the focus of lobster eye optics. Photons which are scattered back toward the detector allow for imaging different materials in the object, since the probability of scattering depends on the number of electrons and then on the atomic number. Such an application requires hard X-rays, which are quite penetrating and interact with matter predominantly with Compton scattering. Moreover, the use of a lobster eye telescope allows to collect enough photons in a reasonable time without the use of a cumbersome grazing incident multilayer optics.

At even larger energies, the interactions produce high-energy electrons that are highly penetrating. In experiments active above 50 MeV photons are converted on a thin layer of a high Z material (W or Pb) into an electron-positron pair. These two particles propagate through a tower of tracking silicon detectors. From the analysis of tracks, the original vertex of interaction is reconstructed and the direction of the original photon is derived. Eventually the pair is absorbed on a thick scintillator to measure the energy.

The Earth from outside: X- and gamma rays from atmosphere

The atmosphere of the Earth seen from space is an intense source of X-rays and gamma rays. Also an intense flux of X-rays and γ -rays falls on the atmosphere from the space: Most of them come from the cosmic diffuse background plus many cosmic X-ray sources, and for the half of the Earth exposed to the Sun, they also arrive from solar flares and steady coronal emission. Some of these photons are absorbed by photoelectric effect, while at higher energies the dominant interaction is the Compton effect. A fraction of photons, of the order of 10 % around 80 keV and decreasing at lower and higher energies, is backscattered to space with a significant shift toward longer wavelengths. But the largest component of photons emitted by the atmosphere derives from the interactions with cosmic rays, which are energetic particles coming from a number of classes of astrophysical sources. Therefore, even though atmospheric emission is usually named “albedo” also in X- and gamma rays, the meaning of this term is quite different from the usual optical definition because actually photons emitted are, for the largest part, not reflected.

Cosmic rays have a large flux with a spectrum extending up to very high energies but quickly decreasing at increasing energies. They are composed mainly of protons (i.e., nuclei of hydrogen for ~ 89 %), α -particles (i.e., nuclei of helium for ~ 10 %) plus nuclei of higher atomic number elements (for the lasting ~ 1 %). In paths of the order of few tens of grams per square centimeter, which correspond by dividing by the density of the upper layers of the atmosphere to a height of a few tens of kilometers, these particles collide with the protons or the neutrons in the nuclei of atmosphere constituents, losing a part of their energy and producing short lifetime particles named pions. The neutral pions π^0 decay into two photons, while charged pions π^+ and π^- decay into μ^+ and μ^- mesons, that further decay into electrons and positrons, plus a bunch of elusive neutrinos, that have a very low probability to further interact. Due to conservation of momentum, all these products (photons and electrons) have still directions close to the original direction of the cosmic ray particle. The particles and the photons produced after the first interactions have sufficient energy to interact again and generate further particles, giving rise to a so-called shower. We have, therefore, a bunch of photons

and particles moving downward and continuously interacting, increasing in number and decreasing in energy. After crossing a certain amount of atmosphere, the components of the shower lose energy by photoelectric absorption and by ionization, and the number of particles/photons starts to decrease. Eventually only a few μ mesons arrive to the ground. Also, as long as the shower develops, the angular distribution increases and some photons are backscattered. While most of the energy of the shower is lost in the atmosphere, a certain fraction of photons, nuclei, and electrons move upward and succeed to escape. This is the main constituent of the so-called albedo. For satellites in low Earth orbit, between 400 and 2,000 km, this is a significant source of instrumental background. In the past, many semiempirical formulas have been given to describe the albedo. More recently, given the presence of excellent codes to simulate all these interactions and the good knowledge of the Cosmic ray spectra, detailed computations have been performed of the X-/ γ -ray fluxes at different depths. Sazonov et al. (2007) derive from its simulations that the spectrum of X-/ γ -ray albedo from the atmosphere in the range 25–300 keV is well fitted by the function:

$$\frac{dN_\gamma}{dE} = \frac{C}{\left(\frac{E}{44 \text{ keV}}\right)^{-5} + \left(\frac{E}{44 \text{ keV}}\right)^{1.4}} \text{ photons/keV} \quad (6)$$

The spectrum peaks around 50–60 keV. At lower energies we have a fast drop, due to the fact that at lower energies photons are absorbed, and the emission involves thinner and more external layers of the atmosphere. When compared with the sky, the Earth as seen from the space is very bright in γ -rays, while in soft X-rays is dark. The constant C in Equation 6 depends on the magnetic latitude, on the phase of the solar cycle and on the zenith angle. Formulas to compute this normalization can be found in Sazonov et al. (2007). At the peak around ~ 60 keV, the flux integrated on all directions is of the order of 4×10^{-3} photons/(cm² s keV) at the magnetic equator and around six times higher at a magnetic latitude of 65°. Also some nuclear gamma ray lines in the MeV range are present, but they are a minor fraction of the total flux.

The remote sensing of the Earth from space in the hard X/soft gamma band is mainly probing the interaction of Cosmic Rays with the upper layers of the atmosphere. This can identify a method to study the atmosphere of other planets and has been actually used with some interplanetary probe.

A particular case is that of terrestrial gamma-ray flashes (TGFs). They were discovered by the Burst and Transient Source Experiment aboard the Compton Gamma Ray Observatory (Fishman et al., 1994) and are very short bursts of γ -rays, lasting from 0.5 to 4 ms, arriving from the direction of Earth and characterized by a much harder spectrum than typical celestial gamma-ray burst. The very short duration of these events puts a requirement on time resolution and on triggering capability of the instrument, which is usually satisfied by modern satellites. More

events were reported by the Reuven Ramaty High Energy Solar Spectroscopic Imager (Smith et al., 2005) and AGILE arrived to detect a TGF event with a maximum energy of 43 MeV (Marisaldi et al., 2010). The geographic distribution of TGF is peculiar: There is a concentration at very low latitudes, with a higher frequency above the dry land than above the oceans.

While the fact that TGFs are originated within the atmosphere seems out of question, a large debate is in progress about their nature. There is a certain similarity with X-/ γ -radiation observed on ground from lightning. Dwyer (2008) proposes that the energetic seed particle production most likely involves either relativistic feedback or runaway electron production in the strong electric fields associated with lightning leaders or streamers. The phenomenon would originate at relatively low levels in the atmosphere (10–20 km). The Atmosphere Space Interaction Monitor, aboard the International Space Station, will study TGFs, trying to connect them with storms and with the elusive Transient Luminous Events. ASIM includes a coded mask and an array of Hard X-ray detectors to better determine the location of the phenomenon (Neubert et al., 2006).

Remote sensing of solar system bodies: planets, satellites, and minor bodies

If the remote sensing (RS) of the Earth is limited to few topics that do not include the solid or liquid surfaces, the situation is different for planets or other bodies of the Solar System without an atmosphere or with one thin enough to be transparent to X- and gamma rays. Actually experiments of RS have been embarked aboard missions flying by and orbiting around some of these bodies.

In the domain of X-rays, the spectrum of photons from the surface can determine, through the fluorescence lines and the absorption edges, the chemical composition and the abundances of the surface. Depending on the energy, this analysis can probe layers spanning from a few tens of microns for Si to, for instance, 1 mm for nickel. One major problem is that planetary surfaces do not emit X-rays but they can only reprocess those coming from an external source. Usually solar flares are providing such a source but, for an interpretation of fluorescence spectra in terms of quantitative chemical composition, a good knowledge of the impinging spectrum is mandatory. However, this is not very easy to achieve because every solar flare has its own spectrum. In the gamma-ray band, a celestial body can be itself a source of radiation deriving from the intrinsic radioactivity of the rocks or from that produced by the impact of cosmic rays, both primary and solar. Since the cosmic rays are ubiquitous, their characteristic affects less the results than the impinging spectrum in X-rays spectroscopy. The gamma-ray lines provide unique signatures for the presence of elements in a layer of the order of 10 cm.

A gamma-ray spectrometer (GRS) was mounted aboard Apollo 15 and Apollo 16. It was based on a NaI(Tl)

scintillator read with a photomultiplier sensitive in the 0.065–27.5 MeV range. Orbiting around the Moon, these instruments covered about 20 % of the surface. They had no directionality and a row mapping was achieved simply on the basis of proximity to various regions at various orbital phases. By comparison of data accumulated in orbit with samples of the Moon collected from various missions in different sites and returned to the Earth, these data allowed to assess that anorthositic fragments are the main constituents of the Moon's surface.

The Near Earth Asteroid Rendezvous Shoemaker spacecraft orbited around asteroid 433 EROS from February 14, 2000, and after about a year it landed on the surface. It included an X-ray/gamma-ray spectrometer. The X-ray detector was a set of three proportional counters capable to measure spectra in the 1–10 keV range. The gamma-ray detector was made of an internal scintillator of NaI(Tl) with a shield made of BGO, each read by its own photomultiplier. The X-ray spectrometer detected the X-rays by solar flares backscattered with superimposed absorption edges and fluorescence lines. From a modeling of solar flares spectrum, the composition of surface layers was derived. The gamma-ray spectrometer detected nuclear lines from a few elements. This can be considered the first case of remote sensing of a reasonable completeness.

After this first experience, X-ray and gamma-ray spectrometers were included in planetary probes. SMART-1, an ESA mission that orbited around the Moon from 2003 to 2006, included the Demonstration of a Compact Imaging X-ray Spectrometer (D-CIXS), an X-ray spectrometer based on silicon detectors with a narrow field collimator (Grande et al., 2007). An important improvement is the presence of a detector of solar flares (Huovelin et al., 2010), providing the input spectrum needed to convert the measured spectra into abundance of elements. A similar instrument, Chandrayaan-1 X-ray Spectrometer (CIXS) was launched on October 22, 2008, on India's Chandrayaan-1 mission to the Moon (Swinyard et al., 2009).

Another X-ray spectrometer (XRS) and gamma-ray spectrometer (GRS) were aboard the Japanese satellite Kaguya/SELENE operative in 2007–2008 (Hasebe et al., 2008; Okada et al., 2009). The spectral quality was much better than previous instruments because it used CCDs for XRS (combined with a solar flare monitor) and a high-resolution germanium detector for the GRS.

MESSENGER is a mission inserted in orbit around Mercury in 2011 (Schlemm et al., 2007) and, since then, mapping elemental abundances in the surface of the planet (Nittler et al., 2011). The X-ray spectrometer onboard is based on silicon detectors set with a collimator to detect the X-rays from the Sun and another set to detect those reflected by Mercury. A germanium detector is dedicated to gamma rays and neutrons (Goldsten et al., 2007).

DAWN mission to Vesta and Ceres includes a gamma-ray and neutron detector to detect gamma-ray lines (Prettyman et al., 2004). A first map of elemental abundances of Vesta minerals, acquired with low altitude orbits

was published (Prettyman et al., 2012). So far we only described experiments that measure X-ray and gamma-ray spectra, in some case with a field limiter, but never imaging. The first imager devoted to remote sensing will be hosted aboard the Mercury Planetary Orbiter of the BepiColombo mission (Fraser et al., 2009). The mercury imaging X-ray spectrometer uses two focusing instruments composed of two identical silicon active pixel sensors and two X-ray optics based on microchannel plates. One telescope is a lobster eye lens, while the other is in a Wolter type I configuration. With this instrument X-ray remote sensing would significantly decrease the gap with respect to remote sensing in other wavelengths.

Summary

X-rays and gamma rays cover the short wavelength side of the electromagnetic spectrum. Their more frequent interactions with matter are the photoelectric absorption, the Compton scattering, and the pair production, whose relative probability depends on the energy of the photon and on the atomic number of the material. X-rays and gamma rays are detected with proportional counters, semiconductor detectors, scintillators or bolometers. X-rays can be reflected at grazing angles by very smooth surfaces. This allows for the manufacture of telescopes that can arrive to good angular resolutions on very small fields of view. To perform imaging on large field of view particular telescopes, named lobster eye, can be used. At higher energies and for wide field imaging systems combining a mask designed with a particular code and position-sensitive detectors are used.

Our Earth observed from the space is a bright source of X-rays and gamma rays, mainly originated by the interaction of cosmic rays with the top layers of the atmosphere. Therefore, remote sensing in these bands is mainly probing this region. On the contrary remote sensing in X-ray and gamma-ray band is performed on bodies of the Solar System without atmosphere (the Moon, Mercury, asteroids) or with a thin one (Mars), with instruments aboard flying-by, orbiting, or landing probes. This technique is expected to contribute significantly to the study of chemical composition of planetary surfaces.

Bibliography

- Angel, J. R. P., 1979. Lobster eyes as X-ray telescopes. *The Astrophysical Journal*, **233**, 364.
- Berger, M. J., et al., XCOM: photon cross sections database. Available from World Wide Web: <http://www.nist.gov/physlab/data/xcom/index.cfm>.
- Chandra X-ray Observatory website. Available from World Wide Web: http://chandra.harvard.edu/press/chandra_specs.html.
- Dwyer, J. R., 2008. Source mechanisms of terrestrial gamma-ray flashes. *Journal of Geophysical Research*, **113**, D10103, doi:10.1029/2007JD009248.
- Fishman, G. J., et al., 1994. Discovery of intense gamma-ray flashes of atmospheric origin. *Science*, **264**, 1313.
- Fraser, G. W., et al., 1993. X-ray focusing using square-pore microchannel plates: first observation of cruxiform image structure. *Nuclear Instruments and Methods in Physics Research*

- Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **324**, 404.
- Fraser, G., et al., 2009. The mercury imaging X-ray spectrometer (MIXS) on bepicolombo. *Planetary and Space Science*, **58**, 79.
- Gertsenshteyn, M., Jansson, T., and Savant, G., 2005. Staring/focusing lobster-eye hard X-Ray imaging for non-astronomical objects. *Proceedings of SPIE*, **5922**, 59220N–59221N.
- Gerward, L., 1999. Paul Villard and his discovery of gamma rays. *Journal Physics in Perspective (PIP)*, **1**, 367.
- Goldsten, J. O., et al., 2007. The MESSENGER gamma-ray and neutron spectrometer. *Space Science Reviews*, **131**, 339.
- Grande, M., et al., 2007. The D-CIXS X-ray spectrometer on the SMART-1 mission to the moon – first results. *Planetary and Space Science*, **55**, 494.
- Harrison, F. A., et al., 2013. The nuclear spectroscopic telescope array (NuSTAR) high-energy X-ray mission. *The Astrophysical Journal*, **770**, 103.
- Hasebe, N., et al., 2008. Gamma-ray spectrometer (GRS) for lunar polar orbiter SELENE. *Earth, Planets and Space*, **60**, 299.
- Henke, B. L., Gullikson, E. M., and Davis, J. C., 1993. X-ray interactions: photoabsorption, scattering, transmission, and reflection at $E = 50\text{--}30000$ eV, $Z = 1\text{--}92$. *Atomic Data and Nuclear Data Tables*, **54**(2), 181–342. Available on-line at http://henke.lbl.gov/optical_constants/.
- Huovelin, J., et al., 2010. Solar intensity X-ray and particle spectrometer (SIXS). *Planetary and Space Science*, **58**, 96.
- Knoll, G. F., 2010. *Radiation Detection and Measurement*. Hoboken, NJ: Wiley.
- Leo, W. R., 1994. *Techniques for Nuclear and Particle Physics Experiments*. Berlin: Springer.
- Lutz, G., 2007. *Semiconductor Radiation Detectors: Device Physics*. Berlin: Springer.
- Marisaldi, M., et al., 2010. Detection of terrestrial gamma ray flashes up to 40 MeV by the AGILE satellite. *Journal of Geophysical Research*, **115**, A00E13, doi:10.1029/2009JA014502.
- NASA's Earth Fact Sheet. <http://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html>.
- Nittler, L. R., et al., 2011. The major-element composition of mercury surface from MESSENGER X-ray spectrometry. *Science*, **333**, 1847.
- Neubert, T., et al., 2006. The atmosphere-space interactions monitor (ASIM) for the International space station. In Gopalswamy, N., and Bhattacharyya, A. (eds.), *Proceedings of the ILWS Workshop, Goa, India*, p. 448.
- Okada, T., et al., 2009. X-Ray Fluorescence Spectrometer (XRS) on Kaguya: current status and results. In *Proceeding of "40th Lunar and Planetary Science Conference (2009)"*, The Woodlands, TX, ID: 1897.
- Prettyman, T. H., et al., 2004. Mapping the elemental composition of Ceres and Vesta: Dawn's gamma ray and neutron detector. *Proceedings of SPIE*, **5660**, 107.
- Prettyman, T. H., et al., 2012. Elemental mapping by Dawn reveals exogenic H in Vesta's Regolith. *Science*, **338**, 242.
- Price, G. J., et al., 2002a. Hard X-ray imaging with microchannel plate optics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **490**, 276.
- Price, G. J., et al., 2002b. X-ray focusing with Wolter microchannel plate optics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **490**, 290.
- Sazonov, S., Churazov, E., Sunyaev, R., and Revnivtsev, M., 2007. Hard X-ray emission of the Earth's atmosphere: Monte Carlo simulations. *Monthly Notices of the Royal Astronomical Society*, **377**(4), 1726–1736.
- Schlemm, C. E., et al., 2007. The X-Ray spectrometer on the MESSENGER spacecraft. *Space Science Reviews*, **131**, 393.
- Shamos, M. H., 1987. *Great Experiments in Physics: Firsthand Accounts from Galileo to Einstein*. New York: Dover.
- Smith, D. M., Lopez, L. I., Lin, R. P., and Barrington-Leigh, C., 2005. Terrestrial gamma-ray flashes observed up to 20 MeV. *Science*, **307**, 1085.
- Swinyard, B. M., et al., 2009. X-ray fluorescence observations of the moon by SMART-1/D-CIXS and the first detection of Ti K α from the lunar surface. *Planetary and Space Science*, **57**, 744.
- Willingale, R., et al., 1998. Hard X-ray imaging with microchannel plate optics. *Experimental Astronomy*, **8**, 281.

Cross-references

Electromagnetic Theory and Wave Propagation
Radiation (Natural) Within the Earth's Environment
Radiation, Electromagnetic
Radiation, Galactic, and Cosmic Background
Severe Storms

GEODESY*

Calvin Klatt
Geodetic Survey Division, Natural Resources Canada,
Ottawa, ON, Canada

Definition

Geodesy. The branch of Earth Sciences concerned with the size, shape, orientation, and gravity field of the Earth.
Global Navigation Satellite Systems (GNSS). Satellite navigation systems that provide geo-spatial positioning with global coverage.
Very Long Baseline Interferometry (VLBI). A radio-astronomical technique utilizing globally distributed antennas having independent atomic clocks. The antennas collectively form an interferometer or synthetic aperture.
Gravimetry. That branch of the geosciences concerned with the measurement of the force or acceleration of gravity.
International Terrestrial Reference Frame (ITRF). The internationally agreed terrestrial reference frame, defining longitude and latitude at a number of reference points, thus forming the basis for much global positioning.

Introduction

Geodesy is comprised of three “pillars”: the shape of the Earth, its orientation, and its gravity field.

Geodesists first consider the global scale and then proceed to the local. Regional mapping, much geophysics, and positioning with the Global Positioning System (GPS) require first that global aspects be well understood. This has driven geodesy to become perhaps the most international of the sciences.

International services and projects, organized under the auspices of the International Association of Geodesy (IAG) (<http://www.iag-aig.org/>), are needed to coordinate work which cannot be conducted on national or

*Her Majesty the Queen in Right of Canada

continental scales. Technique-specific services today include the International GNSS Service (IGS) (<http://www.igs.org/>), the International VLBI Service for Geodesy and Astrometry (IVS) (<http://www.ivscc.gsfc.nasa.gov/>), the International DORIS Service (IDS) (<http://ids.cls.fr/>), and the International Laser Ranging Service (ILRS) (<http://www.ilrs.gsfc.nasa.gov/>). The International Earth Rotation and Reference Systems Service (IERS) (<http://www.iers.org/>) is responsible for producing products based on information from these services. Products include reference frames and Earth's orientation information. The IAG International Gravity Field Service (IGFS) (<http://www.igfs.net/>) coordinates activities related to the Earth's changing gravity field.

The Global Geodetic Observing System (GGOS) (<http://www.ggos.org/>) is an IAG effort to further integrate international geodetic activities. GGOS contributes to the Global Earth Observing System of Systems (GEOSS) (<http://www.earthobservations.org/>) and is consistent with the Integrated Global Observing Strategy (IGOS) (<http://www.igospartners.org/>) (see *Global Earth Observation System of Systems (GEOSS)*).

The shape of the earth and reference frames

The Earth's shape is approximately that of an oblate spheroid with an equatorial radius of about 6,378 km. This radius is approximately 0.3 % greater than that at the poles, or 21 km. The topography varies from about +9 km (Mount Everest) to −11 km (Mariana Trench). The relative size of these variations accounts for the appearance of the Earth from space as a nearly perfect sphere.

For geodesists, the shape of the Earth is described by the location and velocity of reference points in a geocentric terrestrial Earth-fixed reference frame, called the International Terrestrial Reference Frame (ITRF). Locations may be given in geocentric X, Y, and Z coordinates, or in terms of latitude, longitude, and height above an ellipsoid that approximates the Earth's shape.

The ITRF and the International Terrestrial Reference System (ITRS) are products of the IERS. The ITRS defines the frame origin, scale, orientation, and time evolution. The ITRF is a practical realization of the ITRS: three-dimensional coordinates and velocities of reference points on the surface of the Earth. The ITRF is updated periodically as more observational data is obtained or with improvements in theory or analysis.

The IERS publishes conventions to be followed by contributors, providing agreed numerical standards, reference frame information, geophysical models, and models for propagation of observables (McCarthy and Petit, 2003).

The ITRF approximates a no-net-rotation surface, resulting from an averaging of tectonic and other motions on the Earth's crust. As a result, the velocity of a point in the ITRF will be the sum of that due to the motion of the local tectonic plate, intraplate motion/deformation, and other local effects relative to a global mean of motions.

The ITRF is produced from the combination of data from a number of different techniques. Of interest to users are the low cost and easily deployed Global Navigation Satellite System (GNSS) receivers, dominated today (2008) by the Global Positioning System (GPS). There is more than one GNSS constellation. Apart from the operational US GPS system, Russia's GLONASS is near full operational capability, the upcoming European GALILEO system holds much promise, and China may extend the regional Beidou/Compass system to provide global coverage. Regional Navigation Satellite Systems have been proposed by Japan and India.

GNSS augmentation is the use of additional information to improve that broadcast by the satellites. Augmentation can occur in real time or in postprocessing. A trade-off has always existed between latency and accuracy, but improvements in both are occurring at a very rapid pace. User-obtained accuracy is dependent on many factors, including mode of operation (static or dynamic), single or dual frequency observations, length of observation, and equipment used. Postprocessing of data sets (from anywhere on Earth) can now be done on the internet to cm-level accuracy with a delay of only 90 min after data collection (Mireault et al., 2008). The Wide Area Augmentation System (WAAS) is a real-time augmentation system operated by the US Federal Aviation Administration that provides accuracy at nearly the 1 m level. WAAS is available in the Continental United States as well as most of Canada and Alaska. Similar systems exist in much of the world.

The primary benefit of global reference frames is in navigation, positioning, and Earth science studies. For example, the driving force behind the development of the GPS constellation was the military need for precise positioning anywhere on Earth. Use of global reference frame information benefits users who integrate georeferenced data collected by different people for many different purposes at different times. Related fields include mapping, surveying, and hydrography (see *Geological Mapping Using Earth's Magnetic Field*).

In remote sensing, georeferencing may be performed through occupation with GNSS receivers of identifiable points on imagery. With continued improvement of Earth Observation resolution, accurate orthorectification is increasingly important. Additionally, remote sensing satellites or aircraft may be dynamically positioned via GNSS. Augmentation to decimeter-level precision of kinematic GPS data collected during airborne remote sensing can be easily performed (Mireault et al., 2008). Precise global reference frames are required for many scientific efforts, including global sea-level rise studies and satellite laser altimetry missions.

As precision in geodesy has advanced, the ability to observe contemporary change has arisen (Chao, 2003). Networks to monitor motion are established near tectonic plate boundaries, in regions experiencing glacial isostatic adjustment (postglacial rebound) or other regions of interest. Regional, national, continental, and global networks

with varying data latency, precision, and spatial density exist to respond to geoscientific needs (see [Subsidence](#)).

Geodetic reference frame analyses require accurate modeling of many geophysical effects which affect the observables or instruments, including tidal effects, atmospheric loading, and antenna offset/deformation (McCarthy and Petit, 2003).

Effects on geodetic signals (electromagnetic radiation), experienced while traversing the Earth's ionosphere and troposphere, are very significant. Ionospheric effects can be precisely quantified by observation of multiple frequencies since the refractive index is a strong function of frequency. Taking advantage of this, GNSS systems are being deployed as ionospheric sensors (see [Ionospheric Effects on the Propagation of Electromagnetic Waves](#)).

Hydrostatic delays and those due to water vapor in the troposphere corrupt observables in a manner that is difficult to correct for due to the chaotic nature of the weather. Extended data collection and sensitive analysis are necessary to separate these effects from the desired geodetic information (see [Water Vapor](#)).

GNSS systems are deployed in meteorology in two ways. First, the analysis of fixed continuously operating GNSS systems yields information on atmospheric pressure (hydrostatic delays), including water vapor content above the antennae. Second, limb sounding of GNSS satellite signals from other satellites in lower orbit provides information on atmospheric density profiles (see [Limb Sounding, Atmospheric](#)).

Earth's orientation

The orientation of the Earth is given by the relationship between the terrestrial reference frame and another representing inertial space. This relationship involves the rotation of the Earth, polar motion (geographic position of the rotation axis of the Earth), and the coordinates of the rotation axis in the celestial sphere ("celestial pole"). Each of these Earth orientation components exhibits some degree of motion that is currently unpredictable.

Polar motion includes a drift of about 12 m per century superimposed on the Chandler Wobble (having a diameter of about 15 m and a period of approximately 433 days) and a smaller seasonal signal of 8 m diameter. The slow drift of the pole is due to large-scale mass redistribution such as postglacial rebound.

Motion of the celestial pole is the sum of precession (one cycle each 25,800 years) and nutation (a series of many shorter period terms).

Following the invention of accurate clocks based on crystal oscillators during the Second World War, it soon became clear that the Earth's rotation was irregular. Instrumentation was in place to measure the changing Earth rotation prior to the introduction of conventional Universal Time in 1960. The rotation rate may be expressed as its inverse, the length of day.

The Earth system ("solid Earth," oceans, atmosphere) approximates a closed system which conserves angular

momentum. Thus, the counterpoint to variations in Earth's rotation/orientation ("solid Earth") can be seen in atmospheric and oceanic changes. The Earth's orientation thus provides a global sensor of mass movements both within (the "solid" Earth is not solid) and outside the Earth.

The IERS publishes Earth Orientation Parameters which relate the ITRF with the International Celestial Reference Frame (ICRF). The ICRF approximates an inertial reference frame and is primarily expressed by a set of coordinates of quasars, extragalactic objects at great distance from the Earth. This great distance ensures that quasars show no apparent movement (unlike visible stars) and are thus nearly ideal points of reference. The astronomical Very Long Baseline Interferometry (VLBI) technique is critical to establish the ICRF and its link to the ITRF (i.e., the Earth's orientation). Satellite systems are crucial in establishing the origin of the ITRF.

The historical link between time and longitude continues today in the use of GNSS systems as time references (telecommunications) and for precise time transfer, most notably among timing laboratories used in Universal Time definition.

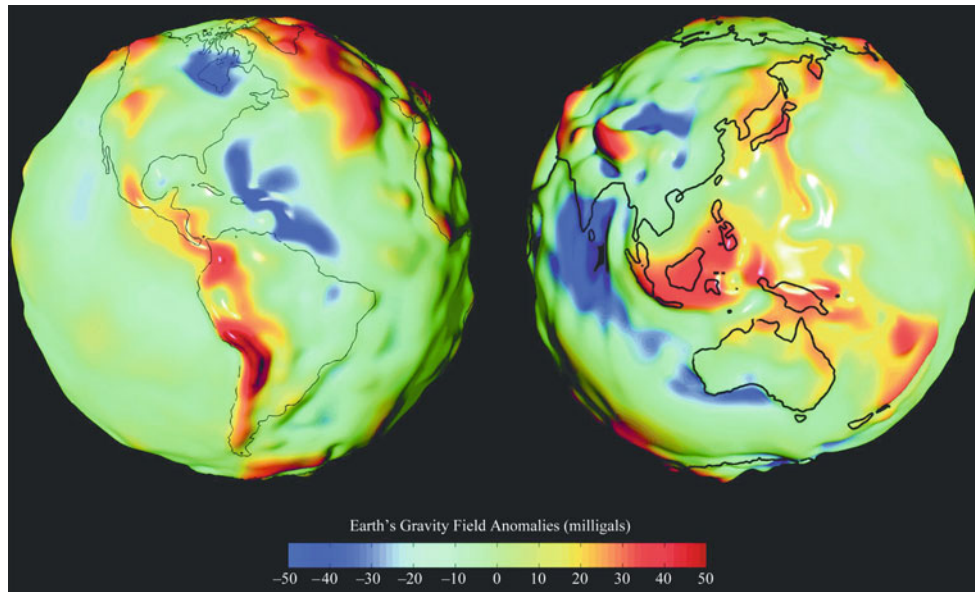
Physical geodesy

The study of the gravity field of the Earth is known as physical geodesy. The force of gravity varies at the surface of the Earth due to distance from the geocenter, density variations in the crust, and other effects.

The Earth is neither homogeneous nor smooth, so the gravitational potential at its surface varies spatially. To describe these variations, we define an equipotential surface (the "geoid") which corresponds to the mean ocean surface of the Earth ("mean sea level"). The geoid is typically described by a smooth ellipsoid plus local "geoid undulations" which can be as large as 107 m (NIMA World Geodetic System Technical Report 2000) ([Figure 1](#)).

Heights on mapping products are given as heights above sea level; thus, they are representative of the inhomogeneous Earth and are useful for management of water. Heights have traditionally been obtained by sequential spirit level measurements from a reference point (e.g., tide gauge), but today it is possible in many regions to use GNSS and a geoid model. Geoid models are used with measurements of sea-surface height to identify sea-surface topography associated with currents (see [Sea Surface Wind/Stress Vector](#)).

Gravity field information contributes to petroleum and mineral exploration, as mineral deposits will have characteristic gravitational signatures. Similarly a characteristic gravitational signature occurs where continental crust meets oceanic crust; thus, gravimetry is used to support efforts to define limits to continental shelves, crucial for delimiting national borders under the United Nations Law of the Sea (see [Solid Earth Mass Transport](#)).



Geodesy, Figure 1 The Earth's gravity field derived from GRACE satellite data (Image courtesy of NASA's Gravity Recovery and Climate Experiment, The University of Texas at Austin's Center for Space Research).

Gravimetry (at the surface, airborne, or from space) can be considered a remote sensing tool that sees through solid objects. Gravity combined with other information (magnetic, seismic, surface geology, etc.) is used to produce three-dimensional maps of potential oil- or mineral-bearing regions. New technologies fusing gravity data with other remotely sensed information, such as hyperspectral surface geochemistry or **RADAR** interferometry, have the potential to further contribute.

Models of the Earth's gravity field are also used in satellite orbit determination and to predict missile trajectories.

The Earth's gravity field has a temporal structure as well. An instrument on the Earth's surface will record large diurnal changes resulting from solid Earth tides (on the order of 30 cm), since the instrument will be sensitive to geocentric distance. Gravity measurements on the Earth's surface in regions experiencing glacial isostatic adjustment (postglacial rebound) detect both the rising surface of the Earth above the geocenter and mass redistribution in the mantle.

Other temporal changes in Earth's gravity are associated with hydrology, as these are the largest mass redistributions on a time scale of months to years. Satellite missions such as GRACE (<http://www.csr.utexas.edu/grace/>) are capable of remotely sensing the changing water mass on and in the Earth (Velicogna and Wahr, 2006). Based on observations of the motion of and relative motion between the two GRACE satellites, a gravity field model is determined at regular (monthly) intervals and expressed in spherical harmonic functions. Such satellites are entirely unlike other remote sensing satellites which

use electromagnetic radiation, frequently incorporate CCD sensors, and produce images composed of pixels.

Conclusion

Geodesy is one of the oldest and most international of the sciences. Rapid technological improvements have transformed our capabilities in recent years, enabling exciting new discoveries about the Earth system. It has been argued that there is a space-geodetic "Moore's law" in which we see a tenfold improvement every decade in measurement precision as well as, in some cases, spatial and temporal resolution (Chao, 2003). These advances are opening up new opportunities such as real-time GNSS-based tsunami warning systems and hydrology from satellite gravimetry.

User access to the ITRF is expected to rapidly improve in the near future as the capacity to perform satellite orbit/clock modeling in near-real time becomes a reality and as multiple satellite constellations are integrated. With this will come many challenges as well as opportunities to learn more about the Earth system.

The VLBI community is focused on a new operational model entitled VLBI2010 which promises significant advances in precision with lower cost operations. Similarly the satellite laser ranging (ILRS) community has a modernization effort underway entitled SLR2000. Refer to the respective IGS service websites for additional information.

The GRACE satellites represent a revolutionary new technique for which applications are still being discovered. Next-generation satellites of this type (in the planning stage) will provide significant improvements and will likely lead to many new benefits.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Chao, B., 2003. Geodesy is not just for static measurements any more. *Eos*, **84**(16), 145–150.
<http://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf>.
 McCarthy D. D., Petit G., (eds.), 2003. IERS Conventions. <ftp://tai.bipm.org/iers/conv2003/tn32.pdf>.
 Mireault, Y., Tetreault, P., Lahaye, F., Heroux, P., and Kouba, J., 2008. Online precise point positioning: a new, timely service from natural resources Canada. *GPS World*, **19**(9), 59–64.
 NIMA World Geodetic System Technical Report, Amendment 1, 2000. NIMA TR8350.2.
 Velicogna, I., and Wahr, J., 2006. Acceleration of Greenland ice mass loss in spring 2004. *Nature*, **443**, 329–331.

Cross-references

Geological Mapping Using Earth's Magnetic Field
 Global Earth Observation System of Systems (GEOSS)
 Ionospheric Effects on the Propagation of Electromagnetic Waves
 Limb Sounding, Atmospheric
 Sea Surface Wind/Stress Vector
 Solid Earth Mass Transport
 Subsidence
 Water Vapor

GEOLOGICAL MAPPING USING EARTH'S MAGNETIC FIELD

Vernon H. Singhroy¹ and Mark Pilkington²

¹Applications Development Section, Natural Resources Canada, Canada Centre for Remote Sensing, Ottawa, ON, Canada

²Geological Survey of Canada, Ottawa, ON, Canada

Definition and introduction

The Earth's magnetic field comprises two main components: the core field, caused by motions of the electrically conductive, liquid iron mixture in the outer core, and the crustal field, caused by magnetic minerals within rocks in the Earth's crust. Geological applications focus on the crustal field, which has spatial variations with wavelengths <1,000 km and amplitudes generally <1,000 nT (nanotesla). The separation of core and crustal field is effected through removal of a mathematical model of the core field known as the IGRF (International Geomagnetic Reference Field). The IGRF is based on mainly magnetic observatory, ship- and satellite-based magnetometer measurements and is updated every 5 years. Once isolated, the (crustal) magnetic field can be used to infer the magnetic properties of the rocks within the Earth's crust and hence provide indications of geological structure and composition. Measurements of the Earth's magnetic field for

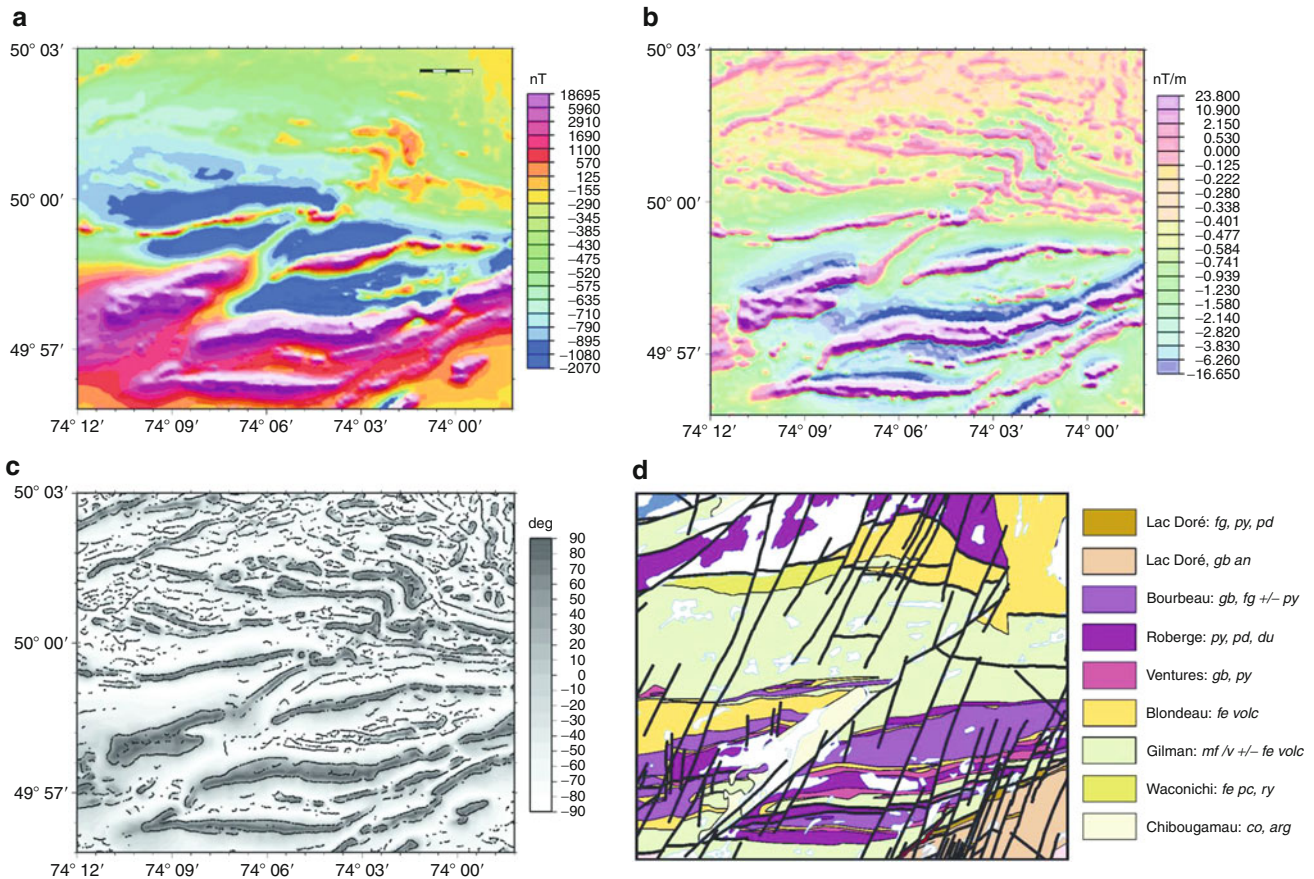
mapping purposes are carried out on a variety of moving platforms: satellites, ships, airplanes, helicopters, and vehicles. Ground-based, handheld magnetometers may also be used for small-scale surveys. The type of measurement platform will be determined by the size of survey area, its topography, location, and the mapping goals. For geological mapping purposes, airborne (fixed-wing aircraft) surveys are the most cost-effective and informative approach for data gathering.

Geological mapping applications of magnetic field data

The fundamental tenet of geological mapping with magnetic fields is *trends of features in the magnetic field mimic those of the outcropping geology*. This is especially the case where the geology consists of magnetic metamorphic and igneous lithologies (Figure 1). Since surficial cover such as vegetation, soil, desert sands, glacial till, water, and many sedimentary rocks is effectively nonmagnetic, the observed magnetic field can be directly related to the underlying rock types. Although structural features and changes in lithology are reflected in the magnetic field maps, a one-to-one connection between some geological feature, e.g., a pluton, and the resulting magnetic field is not always guaranteed. There must be a contrast in the physical property (i.e., magnetization) between the pluton and its surroundings that is large enough to produce a magnetic anomaly which can be detected by a magnetic survey.

The identification of lithology from magnetic field maps can also be ambiguous as the magnetic properties of a given rock type are controlled by the amount and composition of the accessory minerals present, which only constitute about 1.5 % of crustal minerals (Clark and Emerson, 1991). Accessory minerals are generally ignored in petrological classifications (based on silicate mineralogy), and so specific lithologies may exhibit a wide range of magnetic properties. Nevertheless, individual anomalies may signify the presence of a specific lithology or formation. Patterns of anomalies and the shapes of groups of anomalies may reflect certain lithologies, formations, or geological domains. Since there is ambiguity present in what lithology produces a given anomaly, it is desirable to have some knowledge of the magnetizations of the rock types likely encountered in a given area. This knowledge can come from known (published) compilations of physical rock properties, or preferably, from measurements on samples taken from the area of interest or from adjacent areas.

The magnetization of a rock consists of two components: induced and remanent. Induced magnetization is quantified by the rock's susceptibility and is only produced in the presence of the Earth's magnetic field. Remanent magnetization is a permanent magnetization acquired at some point in the rock's history. Both types of magnetization may be present in a rock, but magnetic anomalies are predominantly caused by induced magnetization. The main contributor to the induced magnetization is magnetite (Fe₃O₄), a member of the magnetite-ulvöspinel solid-solution series known as



Geological Mapping Using Earth's Magnetic Field, Figure 1 (a) Magnetic field over portion of Abitibi Greenstone Belt, Quebec, Canada. (b) Magnetic vertical gradient. (c) Tilt of magnetic field with superimposed calculated contact locations. (d) Geological map: *fg* ferrogabbro, *py* pyroxenite, *pd* peridotite, *gb an* gabbroic anorthosite, *gb* gabbro, *du* dunite, *fe volc* felsic volcanics, *mf lv* mafic lavas, felsic pyroclastics, *co* conglomerate, *arg* argillite, *ry* rhyolite. Solid black lines faults.

titanomagnetites (Reynolds et al., 1990). The susceptibility of titanomagnetites is effectively zero for ulvospinel contents greater than about 70 % but is significant and fairly uniform for the rest of the series. Other magnetic minerals of lesser importance are hematite ($\alpha\text{Fe}_2\text{O}_3$), common in oxidized igneous rocks and sediments formed in oxidizing conditions, and pyrrhotite (Fe_7S_8), found in basic igneous rocks, low- to medium-grade metamorphic rocks, and sedimentary rocks.

Interpretation methods

How magnetic maps are interpreted depends on the goal of the study, but the aim is always to characterize and classify the magnetic sources that produce the observed magnetic field. Interpretation methods can be divided broadly into two categories: quantitative and qualitative. The difference between the two is that the former aims to estimate the values of some or all of the physical quantities (e.g., depth, shape, magnetization) of the magnetic sources present, while the latter attempts to classify the

magnetic field (or processed versions of it) into regions that can be interpreted in terms of lithology and structure. For geological mapping applications, qualitative interpretation methods are the most appropriate since the aim is to take the magnetic field data and convert them into a pseudo-geological map. Although automated methods for classifying images are available, most magnetic map interpretations of this type are done manually. Quantitative approaches are usually brought into play afterward, to investigate certain anomalies of interest by, e.g., estimating the dip or depth of a given unit.

For geological mapping applications, the magnetic field map (or image) is not the only piece of information available for interpretation. Numerous processing methods can be applied to the magnetic field data to produce transformed data (or images) where certain features that will aid in the mapping process are preferentially enhanced. Several commonly used transformations are described in the following section.

Magnetic maps

Magnetic field maps

An example of a magnetic field map, covering a portion of the Abitibi Greenstone Belt, Quebec, is presented in [Figure 1a](#). The patterns of colors depicted on the map show the strength of the measured magnetic field and hence reflect how strongly magnetized the rocks are; reds denote magnetic highs and blues represent magnetic lows. Trends in the anomalies reflect the distribution of magnetic minerals (predominantly magnetite) in the surface geology in the area.

The character of individual anomalies is dependent on the shape of the magnetized region or source, its depth, and its magnetization. As the distance from the measurement point to the magnetized source increases, the resulting anomaly decreases in amplitude and increases in width or wavelength. Therefore, the field observed in [Figure 1a](#) is mostly caused by the rocks close to or at the Earth's surface, i.e., the outcropping geology. There will be (generally weaker) contributions from deeper magnetized sources, i.e., buried geological units, but these effects can be considered unimportant when using the data for geological mapping purposes. [Figure 1a](#) shows clearly the predominant east–west striking geological trends in this region (see [Figure 1d](#) for the geology), with the more magnetic units (stronger magnetic field) occurring in the south. Some broad correlations can be made: The highest-amplitude magnetic highs on the map are caused by peridotitic rocks of the Roberge formation. Smaller-amplitude highs are associated with ferrogabbros occurring within anorthosite of the Lac Doré formation in the southeast. Magnetic lows or subdued zones are generally associated with the felsic Blondeau and Waconichi formations. On the other hand, some mafic lithologies do not show a significant magnetic effect, e.g., the Bourbeau and Ventures rocks. Often, more mafic lithologies have more Ti-rich magnetites which results in lower susceptibilities. Within the Gilman formation (volcanics), many magnetic trends can be seen suggesting possible subdivision into mafic and felsic flows. The correlation between geology and magnetic anomalies is made clearer through transformations of the magnetic field as outlined below and shown in [Figure 1b](#) and [c](#).

Magnetic vertical gradient maps

The vertical gradient (or first vertical derivative) of magnetic fields has been used since the 1960s to improve the resolution of features on a magnetic map. The vertical gradient can be measured directly when surveying but is more commonly calculated from the magnetic field. Calculating the vertical gradient is similar to high-pass filtering the data, so long-wavelength anomalies (usually associated with deeper features) are suppressed and gradient anomalies produced by near-surface geological features are emphasized. Closely spaced geological units are better resolved because the vertical gradient of a given anomaly is narrower than its magnetic field

counterpart. Consequently, the vertical gradient map is a standard product used in mapping. As with many enhancement procedures, the problem of noise amplification is present when gradients are calculated; thus, some filtering (usually low-pass) may be required.

[Figure 1b](#) shows the magnetic vertical gradient over the study area. Note that the high-amplitude E–W trending anomalies in the southern part of the map are now better resolved; the anomaly widths have decreased and in some cases a single magnetic field anomaly in [Figure 1a](#) is now resolved into two separate gradient anomalies (e.g., at 74° 09'W 49° 58'N).

Tilt maps

The tilt is defined as the arctangent of the ratio of the vertical gradient to the magnitude of horizontal gradient of the magnetic field (Miller and Singh, 1994). The horizontal gradient achieves a maximum near or over source edges and tends to zero elsewhere. The vertical gradient is positive over the source, zero over or near source edges, and negative elsewhere. As a result, the tilt of the magnetic field tends to be positive over sources bodies, zero over source edges, and negative elsewhere. One advantage of the use of the tilt is that, being a ratio, anomalies due to both weak and strongly magnetized sources are given equal weight. Hence, subtle (and often short-wavelength) anomalies are enhanced. Linear and quasi-linear geological features such as faults appear as minima, whereas dykes often appear as maxima in the tilt map.

[Figure 1c](#) shows the tilt of the magnetic field over the study area as a grey shade map (the superimposed dots are discussed below). Note that the broad change from high to low magnetic field values going northward ([Figure 1a](#)) is not apparent on the tilt map. Also, the prominence of high-amplitude (in the south) versus low-amplitude (in the north) anomalies is removed, and all anomalies have similar expressions (amplitudes) in the tilt map. Both of these effects are due to the balancing effect of calculating a ratio of quantities based on the magnetic field.

Contact maps

Mapping the locations of lateral magnetization contrasts, i.e., the edges of magnetic sources, is one of the most basic and useful applications of magnetic survey data in geological mapping. Delineating the extent of units with similar magnetic properties follows the geological mapping approach of dividing the surface into rocks having similar properties. Whether or not the magnetization changes exactly mimic lithological changes, mapping lateral contrasts provides invaluable information on structural regimes, and deformation styles and trends. There is more than one method of contact mapping, each with its own pros and cons. All methods aim to define contact locations based on maximizing a given function over the source body edge. The locations of these maxima can then be plotted over the magnetic field, vertical derivative, etc., to enhance the interpretation. Details on how the different methods are implemented are given in Pilkington and

Keating (2004). No one method has all of the desirable properties for a reliable contact mapper; hence, using more than one approach is always recommended since colocated solutions from different methods provide increased confidence in the reliability of a given contact location and lessen the adverse effects of source magnetization direction, geological dip, and depth extent.

Figure 1c shows contact locations as black dots superimposed over the tilt of the magnetic field. These contacts were computed from the absolute value of the horizontal gradients of the magnetic field in the N–S and E–W directions (Pilkington and Keating, 2004).

Analytic signal maps

The analytic signal is defined as the absolute value of the square root of the sum of the two horizontal and the vertical derivatives of the total magnetic field (Roest et al., 1992). The main property of the analytic signal is it reaches a maximum value directly over lateral magnetization contrasts (contacts) and that this maximum's location is independent of the orientation of the Earth's magnetic field and of the magnetization direction of the bodies. The benefit of these properties is that no assumptions about the nature of the magnetization of the sources are required. However, this is only true for 2D structures. For 3D structures (e.g., intrusives) the shape of the analytic signal is almost but not completely independent of the Earth's field and magnetization directions. The analytic signal gives maxima directly over contacts and sheets (thin dykes). For a thick dyke, two series of maxima slightly offset outward from their true location are observed. As a result, dykes can appear a bit thicker than they really are. The use of derivatives to define the analytic signal accentuates near-surface effects and noise in the data. As a result, the analytic signal maxima often form discontinuous trends, even along seemingly continuous total field anomalies. Nevertheless, the location of analytic signal maxima can be taken to indicate the position of contrasts between differently magnetized bodies and, as such, can be used to locate geological contacts (cf., Pilkington and Keating, 2004).

Shaded relief magnetic maps

Sometimes the intensity and spatial distribution of magnetic variations within a region are such that a magnetic field image may appear somewhat muted, containing little information. Such images can be enhanced using a technique known as shading or shaded relief. The process simulates the effects of light directed on the three-dimensional surface defined by the grid of magnetic anomaly values. The intensity of light reflected back from a slope on the surface will vary according to the steepness of the slope. The proportion of illuminating light reflected back is known as the reflectance and attains a maximum for slopes that are perpendicular to the direction of the light. The overall effect is one of areas of varying brightness and areas of shadow, an effect very

similar to that created by the sun shining over a range of hills. The light direction and its inclination from the horizontal can be varied, thereby providing different images of the same data set. Depending on the direction and inclination, certain features may be enhanced or suppressed. Generally, magnetic features are enhanced when the light direction is oriented at right angles to them.

Magnetic and radar fused image

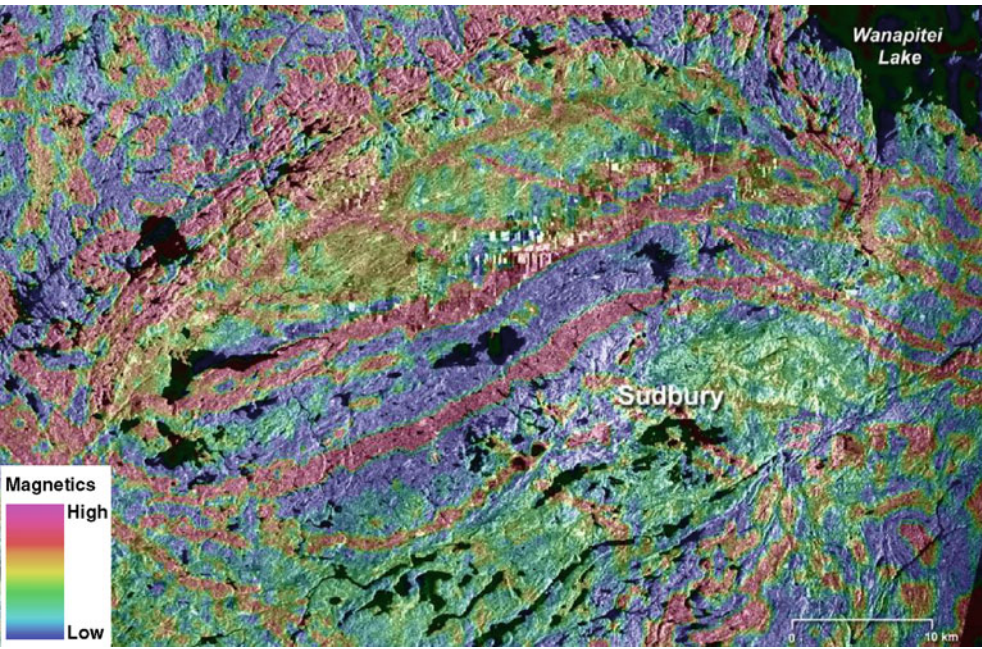
The fusion of magnetic and radar images has been used to assist in geological mapping and mineral exploration (Singhroy and Molch, 2004). An example of the uses of image fusion techniques is presented.

The Sudbury Basin is one of Canada's richest mining areas, with world-class mineral deposits and the world's oldest (2 billion years), largest, and best-exposed meteorite impact structure. Figure 2 shows the 300 km diameter elliptical impact structure known as the Sudbury. Over the past 100 years, about \$135 billion of nickel and copper ores have been mined from more than 90 mines distributed around the rim of the Sudbury Basin. Current production is about \$2 billion a year.

Much remains to be learned about the structural evolution of the Sudbury Basin and the mineral occurrences around it. The RADARSAT-1 image provides an excellent synoptic view of the topography and structural features of the Sudbury area. Structure is expressed by differential erosion controlled by faults, dykes, and different rock types. The fused RADARSAT and magnetic images shown in Figure 2 are produced from special image processing techniques using the Intensity-Hue-Saturation (IHS) integration procedure (Singhroy and Molch, 2004). The RADARSAT-1 image provides terrain information and the surface expressions of structures seen on the image, which appear as a network of linear features. The low-resolution vertical magnetic gradient image (100 m), obtained from the Geological Survey of Canada, shows geological units as expressed from their magnetic signatures. The variation in the concentration of magnetic minerals can be used to map different rock types. Therefore, the combined RADARSAT-1 and the magnetic image is useful in the geological mapping of structures and rock types. The circular area in red (Figure 2) represents the high magnetic signature of the Sudbury Igneous complex. Magma-filled dykes representing NE–SW trending lineaments also have a high magnetic signature. The other colors shown on the fused image correspond to different rock types with different magnetic signatures. These type of fused image maps are used by geologists and mining companies to facilitate geological mapping and mineral exploration projects.

Interpreting magnetic field maps

Geological mapping from magnetic data is essentially an exercise in pattern recognition. The basic approach is to divide the region into areas that have similar character, based on the amplitude, shapes, and textures seen on the magnetic maps. Areas can be initially divided up based on the



Geological Mapping Using Earth’s Magnetic Field, Figure 2 Fusion of RADARSAT and vertical gradient magnetic image of the Sudbury Basin impact structure, Ontario, Canada. This type of image map is used to map geological structures and rock types for mineral exploration (Singhroy and Molch, 2004).

Geological Mapping Using Earth’s Magnetic Field, Table 1 Magnetic field products and their utility for geological mapping applications

Quantity	Use	Mapping application
Magnetic field	General	Lithology + structure
Shaded relief	Emphasize specific directional trends + fine detail	Mainly structure
Vertical magnetic gradient	Emphasize fine detail	Lithology + structure
Tilt	Emphasize low-amplitude areas	Mainly structure
Analytic signal	Highlight contacts, remove magnetic field inclination effects	Mainly contacts
Contact	Map contacts	Structure + contacts

amplitude of the anomalies, i.e., separating the area into regions where amplitudes are high and nonmagnetic regions where amplitudes are low. Negative anomalies on maps do not necessarily imply negative or reversely magnetized rocks; they generally just show areas of weakly magnetized lithologies. Once amplitude information is taken into account, then the patterns/textures of trends/features within these areas can be evaluated. Different “units” can be distinguished based on the strike of anomaly trends by grouping together regions having similar strike (assuming a constant direction is apparent). Similarly, regions can be defined on the basis of similar texture or pattern of anomalies. The resulting areas do not yet have a geological interpretation; they are merely areas defined by similarities in amplitude, trends, and patterns of anomalies. Broad geological inferences can, however, be made. Outlines of the subdivisions may indicate lithology or structure, e.g., ovoid, magnetic regions suggest intrusions, particularly if they interrupt the

surrounding dominant strike direction(s). Sinuous trends suggest ductile deformation, while truncation of trends or region outlines would suggest brittle deformation and faulting. Inferring rock type in a region defined by the magnetic data requires additional independent geological knowledge, usually from geological mapping within or adjacent to the area of interest. In this process, magnetic anomaly characteristics can be compared and calibrated with the known geology and correlations between the two can be established. Importantly, this comparison may dictate which lithologies can be distinguished and which cannot. It is more likely that more than one lithology will produce similar strength anomalies, since magnetic properties may not vary significantly between different rock types, so this ambiguity will remain in the interpreted map.

No one type of magnetic map will provide all the information to make a complete interpretation. Table 1 lists the most useful map products and what their main

applications are. Clearly, there is overlap between the different products, but since they all provide somewhat independent representations of the original data, it is recommended that all that are available be used in an interpretation.

Bibliography

- Clark, D. A., and Emerson, D. W., 1991. Notes on rock magnetization characteristics in applied geophysical studies. *Exploration Geophysics*, **22**, 547–555.
- Miller, H. G., and Singh, V., 1994. Potential field tilt – a new concept for location of potential field sources. *Journal of Applied Geophysics*, **32**, 213–217.
- Pilkington, M., and Keating, P., 2004. Contact mapping from gridded magnetic data – a comparison of techniques. *Exploration Geophysics*, **35**, 306–311.
- Reynolds, R. L., Rosenbaum, J. G., Hudson, M. R., and Fishman, N. S., 1990. Rock magnetism, the distribution of magnetic minerals in the Earth's crust, and aeromagnetic anomalies. In Hanna, W. F. (ed.), *Geologic Applications of Modern Aeromagnetic Surveys*. Denver, CO: U.S. Geological Survey. U.S. Geological Survey Bulletin 1924, pp. 24–45.
- Roest, W. R., Verhoef, J., and Pilkington, M., 1992. Magnetic interpretation using the 3-D analytic signal. *Geophysics*, **57**, 116–125.
- Singhroy, V., and Molch, K., 2004. Geological applications of RADARSAT-2. *Canadian Journal of Remote Sensing*, **30**(6), 893–902.

GEOMORPHOLOGY

David Pieri
Jet Propulsion Laboratory, California Institute of
Technology, Pasadena, CA, USA

Synonyms

Landscape geology; Physical geography

Definitions

Uniformitarianism. The assumption that processes operating at present are the same processes that have operated in the past – essentially that “the present is the key to the past.” The utility of the uniformitarianist paradigm for geomorphology depends on the assumption of a roughly constant range of surface environments, which is not true over on timescales comparable to the age of the planet.

Baselevel. The lowest level to which a stream, river, or groundwater can flow. For large rivers it is usually at the river mouth, typically sea level. For enclosed drainage basins, baselevel can be elevated or “perched.” For most landscapes baselevel limits the depth of erosion.

Pediment. Gently sloping surface of bedrock, often disjoint from steeply sloping mountain foothills, which forms a steady-state equilibrium surface of transport for laminar sheetwash flows. They are often veneered by alluvial gravels or soils.

Scarp. Sometimes also called an escarpment, it is a steep difference in slope, forming a cliff that is often the boundary between different physiographic units within a landscape due to differential susceptibility to erosion. A scarp may also form because of differential uplift along a fault.

Dendrochronology. Also called “tree-ring dating,” is a method of scientific dating of timber to specific years by counting the annual growth patterns of tree rings. As a technique, it has proved effective in the calibration of ages determined by radiocarbon dating over approximately the last 8,000–10,000 years.

Bioturbation. The mixing of sediment grains by the action of plants and animals in the soil.

Introduction

Geomorphology is the study of the configuration and history of the surface of the earth and the processes that shape it. As a scientific discipline, its goals are to understand how and why the landscape has come to look the way it does, through an understanding of dynamic physical and chemical processes through time, and to predict how they will shape its form in the future. Its practitioners, *geomorphologists*, traditionally have approached this challenge with empirical field observations, laboratory and field experiments, and by using numerical and analytical modeling. Modern geomorphic techniques include computer graphics visualization and airborne and orbital remote sensing. While in much of the world, especially the United States, geomorphology is taught and practiced as part of the science of geology, elsewhere, especially in the United Kingdom and former commonwealth countries, its practice and teaching has resided within the field of physical geography. Environmental and civil engineering fields also draw heavily from and contribute to the field. Geomorphology is also often central to the analysis of geological hazards. Most recently, *planetary geomorphologists* have turned their attention to studies of the surfaces of other planets and their satellites as the related field of *comparative planetology* has developed, enabled by spacecraft remote-sensing observations over the last 40 years or more.

While *geomorphological* analyses can be traced back as far as the ancient Chinese literature, the first modern Western geomorphic model was developed by the English pioneer in the field, William Morris Davis (1850–1934), who modeled the cycle of landscape erosion, following the earlier *uniformitarianist* thesis of James Hutton (1726–1797). The classic Davisian model involves riverine downcutting (baselevel lowering), followed by general erosion of the surrounding adjacent landscape to a lower elevation (the peneplain), and followed by uplift to restart the cycle. While conceptually valuable, its gross oversimplification limits its utility as a practical tool. Alternative models by Albrecht Penck (Austria, 1858–1945) and son Walther (Austria, 1888–1923; simultaneous uplift and erosion) and Lester King (South Africa, 1907–1989;

pedimentation via scarp retreat – e.g., King, 1967), competed with the classic Davisian cycle approach, but none were universally applicable. Current geomorphic models embrace the geological paradigm of plate tectonics as one of the underlying forces driving the geomorphic engine as it sculpts the earth. For the other planets, geomorphic processes have played out under a variety of exotic (compared to the earth) boundary conditions, including differing gravitational fields, surface temperature regimes (e.g., cryogenic or hyper-thermal), atmospheric conditions, impact rates, and unique tectonic regimes. G. K. Gilbert (1943–1918), the important early American geomorphologist (dynamic equilibrium model), also was one of the first to apply geomorphology to another planetary surface in his analyses of lunar craters from telescopic observations.

During the so-called quantitative revolution in geomorphology in the middle twentieth century, analyses of drainage networks as a fundamental landscape element were prominent. The fundamental modern contribution to the analyses of drainage networks was made by US Geological Survey hydrologist Robert Horton (1875–1945) in his analyses of the hierarchy of branching stream networks, whereby the smallest unbranched tributaries were assigned order unity, and order added arithmetically downstream throughout the network. Subsequently, Horton's model was modified by the American geologist Arthur Strahler (1918–2002) to include the concept of stream power (the higher the order, the higher the power). Subsequent work by Adrian Scheidegger and Ronald Shreve addressed mathematical inconsistencies in the Horton and Strahler schemes, and Shreve demonstrated the topological randomness of drainage basins, from which much of the systematic character of the stream ordering systems devolved. Time-independent steady-state equilibrium for humid temperate drainage basins, proposed by John Hack (1913–1991), stood somewhat in contrast to the erosion cycle models.

Modern geomorphic work has included an emphasis on quantitative techniques in the field and laboratory and on the physical connection between process and form, and particularly the testing of ideas about the *time-dependent* evolution of landscapes (Ritter et al., 2002). The advent of modern radiometric, dendrochronologic, thermal luminescent, and other absolute dating techniques has made this possible. Tectonic geomorphology has also taken on more importance, particularly over the last 40 years or so as plate tectonics has become the dominant paradigm for earth crustal evolution and dynamics.

Geomorphology of the earth

When seen from space, 75 % of the Earth's surface is dominated by its oceans of liquid water, while the remaining 25 % of nonmarine subaerial land, lying mainly in the Northern Hemisphere, where most of the world's population lives, has been the subject of nearly all historical geological and *geomorphological* study. The Southern

Hemisphere is dominated by oceans, some subaerial continental and archipelago land masses (mainly parts of Africa, South America, Southeast Asia, and Australia), and the large, mainly subglacial, island continent of Antarctica.

Despite 200 years of geological and geographical scientific practice, it has only been during the last 40 years or so that detailed mapping and geophysical explorations of the submarine land surface revealed evidence of “plate tectonics,” namely, the presence of intensely volcanic mid-oceanic ridges and their associated parallel-paired geomagnetic domains, along with submarine trenches along continental or island-arc margins (Heezen and Tharp, 1977). Mid-oceanic ridges are where new volcanically generated material accretes to oceanic plates, where trenches harbor deep subduction, and where oceanic crust is consumed beneath other overriding crustal plates. Less dense continental plates form over time as more silica-rich island-arc material generated volcanically at oceanic plate margins accretes, eventually becoming persistent continental cratons, resistant to subduction because of their lower density, thus having the propensity to “float” over more magnesium and iron-rich oceanic plates. Oceanic plates tend to be mobile and are subducted, thus having typical lifetimes of 100–200 Myr, while continental landmasses may persist up to 2–3 Byr, preserving evidence of much of the history of the earth, including some ancient landscapes. Tectonic plates are the Earth's most fundamental geomorphic units.

Geomorphically, submarine oceanic basins areally dominate terrestrial landforms. The main features of oceanic basins are the oceanic ridge and rise systems. Sometimes locally very rugged, they can rise to several km above the average oceanic depth. In the Atlantic Ocean, rise systems exhibit a central rift valley that is at the center of the rise, which is not always true in the Pacific. Older crust within oceanic basins can have gently rolling abyssal hills. Predominantly flat abyssal plains that stretch for thousands of kilometers are usually also covered with accumulated marine sediment. In places they are punctuated by seamounts – conical topographic rises often topped by coral lagoons or residing just beneath the oceans' surface. These are undersea volcanoes formed as island arcs or midplate hot spots, such as the mid-Pacific Emperor Seamount chain, which terminates in the Hawaiian Islands at its southeastern end. Though areally smaller, oceanic margins are another important submarine landform province. “Atlantic style” continental margins tend to exhibit substantial ancient sediment accumulations and a shelf-slope-rise overall morphology, which probably represents submerged subaerial landscapes remnant from the last Ice Age. Continental shelves are usually less than about 100 km in width and have very shallow ($\sim 0.1^\circ$) topographic slopes. Submarine canyons (also probably remnant from the last Ice Age) can deeply cut the continental shelf and slope and terminate in broad submarine sediment fan deposits at the seaward canyon outlet. “Pacific style” oceanic margins, present along the margins

of the Pacific Rim, are even narrower, consisting of a short shelf and slope that can terminate into deep submarine trenches, including subduction zones (e.g., South America, Kamchatka, Aleutian, and Kurile Islands), up to 10 km deep. Shallower “back-arc” basins occur on the overriding plate behind island arcs (e.g., Sea of Okhotsk).

The subjects of classic geomorphological investigations are the Earth’s “subaerial” landscapes. While mainly continental, some important subaerial landscapes (particularly volcanic ones, e.g., Hawaii, Galapagos Islands) exist on oceanic islands. Terrestrial subaerial landform suites are the classic landscapes studied in geomorphology (Snead, 1980). Drainage basins exist at nearly all spatial scales and range from very active drainage basins in humid and semiarid climatic zones to only occasionally active or relict drainages in arid zones. Drainage basin morphology is strongly influenced by a combination of mountain building related to plate tectonics and prevailing climatic regimes (Schumm, 2005).

Land surface volcanic processes produce characteristic landscapes in all climates. They occur mainly at plate boundaries, with isolated oceanic (e.g., Hawaii) and continental (e.g., the San Francisco volcanic field in Northern Arizona, the Columbia and Snake River volcanic plains in the US Pacific Northwest, the Deccan Traps in India) examples occurring away from plate boundaries. Central vent volcanic landforms range from strato-cone volcanic structures to large collapse and resurgent calderas (e.g., Krakatau, Indonesia). More areally extensive and lower subaerial shield volcanoes, formed by more fluid lavas, exist in places like the Hawaiian Islands or the Galapagos Islands, for example (Francis and Oppenheimer, 2004).

The expenditure of energy in the landscape erodes the earth’s surface and transports sediment. Destructional processes, such as rainfall-driven runoff and streamflow, are essentially exogenic processes (i.e., the energy that drives the evaporation and precipitation, and winds that transport water vapor, comes from an exterior source – the Sun). Constructive landscape processes are mostly endogenic – derived from the internal dynamo processes that drive plate tectonics and subsequent mountain building (Bloom, 1998).

In familiar ways, such destructional geomorphic processes work to reduce the “topographic disequilibria” that constructive landscapes represent. For instance, the relatively low and ancient Appalachian Mountains were pushed up during one of the collisions between the North American and European continental landmasses and probably exhibited relief comparable to the Himalayan mountain range today. The former topographic relief represented strong disequilibria and thus potential energy to be converted to kinetic energy by erosive runoff. Once continental collision processes slowed to a stop and tectonic uplift ceased, erosion and surface transport processes (such as rainfall, associated runoff, snowfall, and glaciation) over only a few tens of millions of years reduced the proto-Appalachian Mountains to the gently sloping and relatively low-relief state we see today. Volcanogenic

landscapes are good examples of how competition between destructive and constructive processes sculpts the earth. For example, Mt. Fuji in Japan is an active volcano that erupts approximately every 100–150 years. Fuji’s classic symmetrical conical shape is the result of volcanic eruptions that deposit material, on the average, faster than it can be eroded. When Fuji enters a dormant phase, it will become deeply incised by stream erosion, and it will rapidly lose its symmetric conical profile over a geologically short interval.

Weathering is another key geomorphic process through which consolidated material is broken down into constituent grains by chemical and physical means. Chemical weathering occurs when natural acids act on carbonates, such as limestone or sandstones, releasing the residual silicate grains. Mechanical weathering occurs when the hydrostatic pressure of ice in freeze-thaw cycles overcomes the brittle strength thresholds in rock at microscopic and macroscopic scales. Likewise, salt crystals can exert mechanical energy to break up rocks and can chemically weather rocks, mostly in arid areas. Mineral oxidation, particularly iron-containing minerals, is yet another form of chemical weathering. Biological weathering can take the form of chemical weathering by biogenic acids, particularly in tropical areas. It can also occur mechanically by bioturbation of soils and sediments, as well as by the physical pressure of root and stem turgor in cracks and fissures within solid rock. It is significant that all three main forms of weathering are enhanced or enabled by the presence of water.

Finally, mass wasting is perhaps the most dramatic form of nonvolcanic landscape alteration. The term “mass wasting” is applied to processes such as landslides; creep, snow, and debris avalanches; submarine slides and slumps; volcano-tectonic sector collapses; and scour related to the action of glaciers. While such processes tend to affect only a small part of the Earth’s surface, when they occur near populated areas, their effects can be devastating.

New technology and global geomorphology

Stimulated by the desire to understand the Earth’s history in the nineteenth century, geomorphology evolved from a highly descriptive natural history to a modern physical science by the progressive adoption of analytical techniques and technologies as they became applicable and available for landscape-related science. New airborne survey techniques developed as a result of World War II provided comprehensive information on the topography and near-subsurface structure and composition of landscapes. Improved aerial photography and photogrammetry greatly impacted quantitative approaches to geomorphology. The Newtonian approaches of Strahler (1950, 1952), Bagnold (1941), Horton (1945), and Leopold and Maddock (1953), emphasizing direct observables, and the advent of absolute dating methods, especially radiocarbon methods (Libby, 1955) for time periods more recent than 30Ka BP, had a profound effect on the field, finally allowing accurate estimations of surface process rates. The

expansion of modern technology within geomorphology also facilitated quantitative statistical approaches to landscape analyses (e.g., Schumm, 1956; Melton, 1958; Lubowe, 1964; Shreve, 1975; and Pieri, 1984) and allowed more complex analyses, such as attempts to illuminate the self-organization of landscapes (e.g., Phillips, 1995, 2003, 2009).

Most recently given the proliferation of large (10–100 Tb) databases produced by a new generation of late twentieth and early twenty-first century orbital observational instruments (e.g., Pieri and Abrams, 2004), high spatial and spectral resolution orbital imaging at wavelengths from visible through thermal infrared to microwave detailed stereophotogrammetric, radar, and LIDAR modeling of land surface topography, and relatively frequent temporal sampling are now available. Modern earth orbital missions routinely allow detailed physical and chemical characterization of the Earth's surface at 10–100 m scales, or better (e.g., ASTER, SRTM, EO-1, Radarsat, ICESAT, and others). The emergence and ready availability of remote-sensing technology and its influence on the science of geomorphology have fostered the modern view geomorphology as a “system science” (Church, 2010) with an integrated global perspective, especially among younger *geomorphologists* who are technologically adept.

These new technologies are now being used to monitor and study the dynamics and history of earth surface processes, land use, and societal vulnerabilities at a variety of scales and to synergistically complement in situ field observations, across the planet. Spatial continuity of data is now available down to the centimeter scale over many large areas (Church, 2010). Global topographic databases (e.g., ASTER Global Digital Elevation Model) with resolutions of tens of meters, and for local areas LIDAR imagery, can provide data at almost granular spatial scales. The availability of such thoroughly quantitative measures of the global landscape has opened up possibilities of analytical analysis that were undreamed of just a few decades ago (Lane et al., 1993). One of the most recent remote-sensing-related technologies that have been of critical importance has been the Global Positioning System (GPS). It has provided unprecedented ability to determine location on the Earth's surface, typically to within meters. Especially when combined with radar interferometry, these techniques allow the dynamic nature of rapidly changing topography (e.g., pre-eruption inflation of volcanoes) to come within view. In the aggregate, this postmillennial remote-sensing and measurement technology bonanza has found important application in volcanology, glacial and periglacial processes, hydrology and fluvial geomorphology, mass wasting analyses, coastal studies, eolian studies, and geomorphology as related to climate change. Perhaps most profoundly, the precision and accuracy of such techniques now allow us to document the impact of man as the principal agent of geomorphic change on Earth (Church, 2010; Hooke, 1994, 2000).

Summary

The Earth had liquid water oceans for most of its history, has a highly mobile crust, and a dynamically convecting interior. Its surface has been constantly driven by the movement of the interior causing the periodic conglomeration and separation of continental landmasses, the opening and closing of oceans, and the construction and destruction of mountain ranges. Such globally dynamic geomorphology profoundly impacted the global climate and thus biological evolution in general, and specifically human habitation patterns and the development of civilization. Given that as a species we have been shaped and affected by (and now, ourselves, shape) the patterns and intensity of surface geological processes – geomorphic processes – it is hard to overemphasize the importance of geomorphology as a discipline that offers environmental insights significant to our continued well-being as Earth inhabitants. New emerging techniques and approaches to landscape analyses using remote sensing on a planetary scale offer additional significant advantages for unification of regional geomorphological insights into an emerging global science (Baker, 1986; Short and Blair, 1986; Church, 2010). Such knowledge will most likely be even more important as we try to cope with and mitigate the effects of global climate change on society.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Bagnold, R. A., 1941. *The physics of blown dunes*. London: Methuen, p. 265.
- Baker, V. R., 1986. Introduction: regional landforms analysis. In Short, N. M., and Blair, R. W., Jr. (eds.), *Geomorphology from Space: A Global Overview of Regional Landforms*. Washington, DC: NASA Scientific and Technical Information Branch, pp. 1–26.
- Bloom, A. L., 1998. *Geomorphology: A Systematic Analysis of Late Cenozoic Landforms*, 3rd edn. Upper Saddle River: Prentice Hall.
- Church, M., 2010. The trajectory of geomorphology. *Progress in Physical Geography*, **34**, 265–286.
- Francis, P. W., and Oppenheimer, C., 2004. *Volcanoes*. Oxford: Oxford University Press.
- Heezen, B., and Sharp, M., 1977. *Panoramic Maps of the Ocean Floor*. Washington, DC: United States Navy.
- Hooke, R. L. B., 1994. On the efficacy of humans as geomorphic agents. *GSA Today*, **4**, 224–225.
- Hooke, R. L. B., 2000. On the history of humans as geomorphic agents. *Geology*, **28**, 843–846.
- Horton, R. E., 1945. Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin*, **56**, 275–370.
- King, L. C., 1967. *Morphology of the Earth*, 2nd edn. Edinburgh: Oliver and Boyd.
- Lane, S. N., Richards, K. S., and Chandler, J. H., 1993. Developments in photogrammetry: the geomorphic potential. *Progress in Physical Geography*, **17**, 306–328.

- Leopold, L. B., and Maddock, T. 1953. The hydraulic geometry of stream channels and some physiographic implications. Reston, VA: United States Geological Survey. Professional Paper 252, 57 pp.
- Libby, W. F., 1955. *Radiocarbon Dating*, 2nd edn. Chicago: University of Chicago Press, p. 175.
- Lubowe, J. K., 1964. Stream junction angles in dendritic drainage pattern. *American Journal of Science*, **262**, 325.
- Melton, M. A., 1958. Geometric properties of mature drainage systems and their representation in an E4 phase space. *Journal of Geology*, **66**, 35–56.
- Phillips, J. D., 1995. Self-organization and landscape development. *Progress in Physical Geography*, **19**, 309–321.
- Phillips, J. D., 2003. Sources of nonlinearity and complexity in geomorphic systems. *Progress in Physical Geography*, **27**, 1–23.
- Phillips, J. D., 2009. Landscape evolution space and the relative importance of geomorphic processes and controls. *Geomorphology*, **109**, 79–85.
- Pieri, D. C., 1984. Junction angles in drainage networks. *Journal of Geophysical Research*, **89**, 6878–6884 (NB8).
- Pieri, D., and Abrams, M., 2004. ASTER watches the world's volcanoes: a new paradigm for volcanological observations from orbit. *Journal of Volcanology and Geothermal Research*, **135** (1–2), 13–28.
- Ritter, D. F., Kochel, R. C., and Miller, J. R., 2002. *Process Geomorphology*, 4th edn. New York: McGraw-Hill.
- Schumm, S. A., 1956. Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. *Geological Society of America Bulletin*, **67**, 597–646.
- Schumm, S. A., 2005. *River Variability and Complexity*. New York: Cambridge University Press, p. 234.
- Short, N. M., and Blair, R. W., Jr. (eds.), 1986. *Geomorphology from Space: A Global Overview of Regional Landforms*. Washington, DC: NASA Scientific and Technical Information Branch. NASA SP-486.
- Shreve, R. L., 1975. The probabilistic-topologic approach to drainage-basin geomorphology. *Geology*, **3**, 527–529.
- Snead, R. E., 1980. *World Atlas of Geomorphic Features*. Huntington/New York: Robert E. Krieger/Van Nostrand Reinhold.
- Strahler, A. N., 1950. Equilibrium theory of erosional slopes approached by frequency distribution analysis. *American Journal of Science*, **248**, 673–696.
- Strahler, A. N., 1952. Dynamic basis of geomorphology. *Geological Society of America Bulletin*, **63**, 923–937.

GEOPHYSICAL RETRIEVAL, FORWARD MODELS IN REMOTE SENSING

Eugene Ustinov
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition

Forward model. Quantitative tool for simulation of *observables* for a given set of *model parameters*.

Forward problem. System of differential equations with initial and/or boundary conditions, solution of which, the *forward solution*, is used for simulation of *observables*.

Forward solution. Solution of the *forward problem*.

Instrument model. Quantitative procedure for evaluation of *observables* from the *forward solution*, which models

the characteristics of the instrument and the procedure for measurements.

Model parameters. Parameters of the forward problem contained in its differential equations and initial and/or boundary conditions.

Observables. Output parameters of the *forward model*, which simulate the quantities measured in the geophysical retrieval.

Sensitivities. Partial or variational derivatives of *observables* with respect to *model parameters*.

Introduction

A general qualitative description of forward models used in remote sensing, their components, and functions is provided in the entry [Geophysical Retrieval, Overview](#). Two major components of the forward model are as follows:

1. *Forward problem* – provides a quantitative description of physical processes in the object of remote sensing, which result into observable quantities used in the geophysical retrieval. In a general case, the forward problem has a form of a system of equations – differential equations with initial and/or boundary conditions. Its solution, the *forward solution*, serves as an input for the procedure of computation of observables.
2. *Instrument model* – an expression converting the forward solution into observables. In a general case, it has a form of a functional, which maps the forward solution, a function, into the observable, a discrete quantity. As such, the instrument model is not a mathematical problem, which has to be solved. It is an expression, which has to be calculated.

Two major functions of the forward model are:

1. *Computation of observables* – involves application of the instrument model to the forward solution to obtain the observables.
2. *Computation of sensitivities* – involves application of the linearized forward problem and linearized instrument model to obtain sensitivities of observables to retrieval parameters.

Implementation of the second function of the forward model can be accomplished in three different ways:

1. Finite-difference approach
2. Linearization approach
3. Adjoint approach

An upper-level quantitative description of major components and functions of the forward model is provided in this entry along with illustrations based on simple analytic demo forward problem and instrument model.

Forward problem, instrument model, and computation of observables

In the general case, the forward problem represents a set of nonlinear differential equations with initial and/or boundary conditions. In high-level notation, the forward

problem can be written in the form of a generally nonlinear operator equation

$$N[X] = S \quad (1)$$

where the operator N combines all operations on the forward solution X , whereas the right-hand term S combines all inhomogeneous terms. All model parameters are contained either in the operator N or in the right-hand term S , which are some functions of these parameters.

As an example, consider a demo forward problem:

$$\begin{cases} \frac{dX}{dt} + a(t)X^2(t) = 0 \\ X(t_0) = X_0 \end{cases} \quad (2)$$

It has two model parameters: $a(t)$ and X_0 . This problem can be written in the form of Eq. 1, where

$$N[X] = \frac{dX}{dt} + a(t)X^2(t) + \delta(t - t_0)X(t)$$

and

$$S(t) = \delta(t - t_0)X_0$$

and the Dirac δ -function enforces action of the initial condition only at $t = t_0$. This forward problem has an analytic solution

$$\begin{aligned} X(t) &= \left(\int_{t_0}^t a(t')dt' + (X_0)^{-1} \right)^{-1} \\ &= \left(A(t) + (X_0)^{-1} \right)^{-1}, \end{aligned} \quad (3)$$

where

$$A(t) = \int_{t_0}^t a(t')dt' \quad (4)$$

is a primitive function of $a(t)$.

Using similar high-level notations, computation of observables R from the forward solution X can be represented in the form

$$R = M[X] \quad (5)$$

where M is, in general, a nonlinear functional describing the procedure of modeling the observables. In many practical applications, this functional is linear, and the above procedure can be represented in the form of the inner product

$$R = (W, X) \quad (6)$$

where the function W describing this procedure has the same domain of arguments as the forward solution X . It has a meaning of a weighting function, “carving” the

observables from the forward solution and is referred to here as *observable weighting function*.

As an example, consider an observable obtained from the solution of the demo forward problem Eq. 2 in the form

$$R = X(t_1) = \int_{t_0}^{t_1} \delta(t - t_1)X(t)dt, \text{ and } W(t) = \delta(t - t_1) \quad (7)$$

Since the forward solution is available in analytic form Eq. 3, dependence of this observable on model parameters $a(t)$ and X_0 is also available analytically:

$$R = \left(A(t_1) + (X_0)^{-1} \right)^{-1} \quad (8)$$

This analytic dependence will be used to illustrate various applications below.

As a more realistic example, consider a forward model of radiances in the thermal spectral region at the top of a plane-parallel non-scattering planetary atmosphere. The forward problem has the form of a radiative transfer (RT) problem:

$$\begin{aligned} u \frac{dI_v}{dz} + \kappa_v(z)I_v(z, u) &= \kappa_v(z)B_v(z), \\ I_v(0, u) &= 0 \text{ for } u > 0, \\ I_v(z_0, u) &= \varepsilon_v B_{s,v} \text{ for } u < 0 \end{aligned} \quad (9)$$

Here, the forward solution $I_v(z, u)$ is monochromatic intensity of radiation at frequency ν depending on vertical and angular coordinates: depth into the atmosphere z and cosine of the nadir angle of propagation of radiation u . The forward problem Eq. 9 has two continuous and two discrete model parameters: atmospheric absorption coefficient $\kappa_v(z)$, Planck function $B_v(z)$, surface emissivity ε_v , and Planck function $B_{s,v}$. The geophysical parameters of interest, the atmospheric and surface temperature, as well as the composition of the atmosphere and surface, are encapsulated in the radiative parameters of the RT problem, and dependence of the radiative parameters on retrieval geophysical parameters is assumed to be known.

The instrument model here is specified by instrument spectral and angular responses. Assuming for simplicity that these responses are linear and decoupled into $\psi(\nu - \nu_0)$ and $\Psi(u - u_0)$, respectively, the radiances measured at the top of the atmosphere ($z = 0$) are represented as

$$R = \int_{\Delta\nu} d\nu \psi(\nu - \nu_0) \int_{-1}^1 du \Psi(u - u_0) \int_0^{z_0} dz \delta(z) I_v(z, u)$$

Thus, observables in this forward model can be represented in the form of a linear functional

$$R = \int_{\Delta\nu} d\nu \int_{-1}^1 du \int_0^{z_0} dz W_\nu(z, u) I_v(z, u) = (W, I)$$

with the observable weighting function

$$W_v(z, u) = \psi(v - v_0)\Psi(u - u_0)\delta(z)$$

Sensitivities to discrete and continuous model parameters: an overview

The dependence of observables on discrete parameters is essentially that of a function of multiple variables. Thus, the sensitivity of the observable R with respect to a discrete parameter p_k is simply a partial derivative $\partial R/\partial p_k$. If a discrete parameter p_k experiences a variation δp_k , then corresponding variation of the observable has the form

$$\delta_{p_k} R = \frac{\partial R}{\partial p_k} \delta p_k \quad (10)$$

The dependence of observables on continuous parameters is essentially that of a functional dependence on a function. Correspondingly, the sensitivities of observables with respect to continuous parameters are variational (aka functional) derivatives, which can be introduced as follows. Let $F[f]$ be a functional defined in the domain of functions $f(x)$, which, in their turn, are defined in the domain Dx of arguments x . By definition, the variational derivative $\delta F/\delta f(x)$ of the functional $F[f]$ with respect to $f(x)$ is a kernel $K(x)$ in the linear integral relation

$$\delta_f F = \int_{Dx} K(x) \delta f(x) dx \quad (11)$$

and correspondingly

$$\frac{\delta F}{\delta f(x)} \equiv K(x) \quad (12)$$

If the function $f(x)$ is approximated by its values $\{f_j\}$ on a finite-dimensional grid of arguments $\{x_j\}$, then the functional $F[f]$ is approximated by a function of multiple variables $F(f_j)$. Correspondingly, variation $\delta_f \tilde{F}$ can be approximated as

$$\delta_f \tilde{F} \approx \sum_j \frac{\partial \tilde{F}}{\partial f_j} \delta f_j \quad (13)$$

On the other hand, the above integral expression for the variation $\delta_f F$ Eq. 11 can be approximated as

$$\delta_f F \approx \sum_j K_j \delta f_j \Delta_j x \quad (14)$$

where

$$K_j = \left. \frac{\delta F}{\delta f(x)} \right|_{x=x_j}$$

This yields an interrelation between the variational derivative $\delta F/\delta f(x)$ and its finite-dimensional approximation $\partial \tilde{F}/\partial f_j$:

$$\left. \frac{\delta F}{\delta f(x)} \right|_{x=x_j} \approx \frac{1}{\Delta_j x} \cdot \frac{\partial \tilde{F}}{\partial f_j}, \text{ and } \frac{\partial \tilde{F}}{\partial f_j} \approx \left. \frac{\delta F}{\delta f(x)} \right|_{x=x_j} \cdot \Delta_j x \quad (15)$$

Note that the value of the function $f(x)$ at some fixed value of the argument $x = x'$ can also be considered as a functional defined on this function. To find the variational derivative $\delta f(x')/\delta f(x)$, we observe that

$$f(x') = \int_{Dx} \delta(x' - x) f(x) dx$$

Comparison with Eqs. 11 and 12 yields

$$\frac{\delta f(x')}{\delta f(x)} = \delta(x' - x) \quad (16)$$

This definition Eqs. 11 and 12 represents a practical recipe used to derive expressions for computation of variational derivatives and, thus, of sensitivities of observables to continuous parameters. If the parameter $p(x)$ experiences a variation $\delta p(x)$ and corresponding variation of the observable R can be represented in the form of Eq. 11

$$\delta_p R = \int_{Dx} K_p(x) \delta p(x) dx \quad (17)$$

then sensitivity $\delta R/\delta p(x)$ is the kernel of this linear integral expression

$$\frac{\delta R}{\delta p(x)} \equiv K_p(x) \quad (18)$$

Using known solution Eq. 3, we can directly obtain sensitivities of observable in the above demo forward model Eq. 2 with respect to model parameters $a(t)$ and X_0 . Sensitivity to the discrete parameter X_0 is obtained as a partial derivative:

$$\frac{\partial R_1}{\partial X_0} = -\left(A(t_1) + (X_0)^{-1}\right)^{-2} \cdot (X_0)^{-2} = \left(\frac{R}{X_0}\right)^2 \quad (19)$$

Sensitivity to the continuous parameter $a(t)$ is obtained using the definition of the variational derivative Eqs. 11 and 12. We have

$$\delta_a R = -\left(A(t) + (X_0)^{-1}\right)^{-2} \int_{t_0}^{t_1} \delta a(t) dt = -R^2 \int_{t_0}^{t_1} \delta a(t) dt$$

Then

$$\frac{\delta R}{\delta a(t)} = -R^2 \quad (20)$$

Sensitivity analysis: computation of sensitivities of observables with respect to model parameters

Finite-difference approach

This approach is based on using differences between values of observables $R = M[X]$ computed from the solution X of the *baseline* (non-perturbed) forward problem $N[X] = S$ and values of observables $\tilde{R}_p = M[\tilde{X}_p]$ computed from the solution $\tilde{X}_p$ of the forward problem $\tilde{N}_p[X] = \tilde{S}_p$ with the operator and right-hand term perturbed due to perturbations $\tilde{p} = p + \Delta p$ of the retrieval parameter. Both components of the forward model, $\tilde{N}_p[X] = \tilde{S}_p$, and $\tilde{R}_p = M[\tilde{X}_p]$ are run separately for each discrete retrieval parameter and for each gridpoint of each continuous parameter. The matrix of sensitivities $K = \partial R / \partial p$ is approximated as

$$K = \frac{\partial R}{\partial p} \approx \frac{\tilde{R}_p - R}{\tilde{p} - p}$$

No analytic work is required here, but computer time, as compared to a single run of the forward model, increases in a direct proportion to the number of (gridpoints of) model parameters of interest.

Linearization approach

In this approach, the forward model Eq. 1 has to be linearized assuming that all model parameters of interest experience infinitesimally small perturbations. The resulting linearized forward problem has the form

$$L\delta X = \delta S - \delta N[X] \quad (21)$$

where the linear operator L represents a linearization of the operator N around the baseline solution of the forward problem X .

Note that only the right-hand term of the linearized forward problem Eq. 21 depends on variations of model parameters; the operator of the linearized forward problem L remains the same. This provides substantial savings of computer time.

After the linearized forward problem is formulated, it is solved for sensitivities of the forward solution $X(x)$ to model parameters: discrete p_j and continuous $p(x')$:

$$L \frac{\partial X}{\partial p_j} = \frac{\partial S}{\partial p_j} - \frac{\partial N[X]}{\partial p_j}, \text{ and } L \frac{\delta X}{\delta p(x')} = \frac{\delta S}{\delta p(x')} - \frac{\delta N[X]}{\delta p(x')}$$

Recalling that both N and S depend on continuous parameters as functions, and using the property Eq. 16, the linearized forward problem for sensitivities of the forward solution to continuous model parameters can be rewritten as

$$L \frac{\delta X}{\delta p(x')} = \delta(x' - x) \left(\frac{\partial S}{\partial p(x)} - \frac{\partial N[X]}{\partial p(x)} \right) \quad (22)$$

Further on, the instrument model Eq. 5 is linearized with respect to the forward solution X , resulting into a linear functional, which is written in the form of an inner product

$$\delta R = (W, \delta X) \quad (23)$$

This yields resulting expressions for sensitivities of observables to discrete and continuous parameters:

$$\frac{\partial R}{\partial p_j} = \left(W, \frac{\partial X}{\partial p_j} \right), \quad \frac{\delta R}{\delta p(x')} = \left(W, \frac{\delta X}{\delta p(x')} \right) \quad (24)$$

As an example, consider the linearization of the demo forward problem Eq. 2:

$$\begin{cases} \left(\frac{d}{dt} + 2a(t)X(t) \right) \delta X(t) = -X^2(t) \delta a(t) \\ \delta X(t_0) = \delta X_0 \end{cases}$$

Here we have

$$\begin{aligned} L\delta X &= \left(\frac{d}{dt} + 2a(t)X(t) + \delta(t - t_0) \right) \delta X(t), \\ \delta S(t) &= \delta(t - t_0) \delta X_0, \quad \delta N[X] = X^2(t) \delta a(t) \end{aligned} \quad (25)$$

We obtain sensitivities to parameters X_0 and $a(t)$ using known analytic solutions of this problem.

For X_0 , corresponding problem has the form

$$\begin{cases} \left(\frac{d}{dt} + 2a(t)X(t) \right)_{\delta X_0} = X(t) = 0 \\ \delta_{X_0} X(t_0) = \delta X_0 \end{cases}$$

Comparing its solution

$$\delta_{X_0} X(t) = \delta X_0 \exp \left(-2 \int_{t_0}^t a(t') X(t') dt' \right)$$

with definition Eq. 10, we have

$$\frac{\partial X(t)}{\partial X_0} = \exp \left(-2 \int_{t_0}^t a(t') X(t') dt' \right)$$

Using $A(t)$, a primitive function of $a(t)$ defined by Eq. 4, and transforming

$$\begin{aligned} \int_{t_0}^t a(t') X(t') dt' &= \int_{t_0}^t X(t') dA(t') \\ &= \ln \left[A(t') + (X_0)^{-1} \right] \Big|_{t_0}^t = \ln \left(\frac{X_0 A(t) + 1}{X_0 A(t_0) + 1} \right) \\ &= \ln \left(\frac{X_0}{X(t)} \right) \end{aligned}$$

yield

$$\frac{\partial X(t_1)}{\partial X_0} = \left(\frac{X(t_1)}{X_0} \right)^2$$

Thus

$$\frac{\partial R}{\partial X_0} = \left(\frac{R}{X_0} \right)^2 \quad (26)$$

in accordance with Eq. 19 obtained above using direct linearization.

For $a(t)$, the solution of the corresponding problem

$$\left(\frac{d}{dt} + 2a(t)X(t) \right) \delta_a X(t) = -X^2(t) \delta a(t), \delta_a X(t_0) = 0$$

has the form

$$\delta_a X(t) = - \int_{t_0}^t dt' \delta a(t') X^2(t') \exp \left(-2 \int_{t_0}^t a(t'') X(t'') dt'' \right)$$

Using $A(t)$, Eq. 4 and transforming

$$\begin{aligned} \int_{t'}^t a(t'') X(t'') dt &= \int_{t'}^t X(t'') dA(t'') \\ &= \ln [A(t'') + (X_0^{-1})] \Big|_{t'}^t = \ln \left(\frac{X_0 A(t) + 1}{X_0 A(t') + 1} \right) = \ln \left(\frac{X(t')}{X(t)} \right) \end{aligned}$$

yields the solution $\delta_a X(t)$ in the form suitable for direct computation of the sensitivity $\delta X(t)/\delta a(t')$:

$$\delta_a X(t) = -X^2(t) \int_{t_0}^t dt' \delta a(t'), \text{ and } \frac{\delta X(t)}{\delta a(t')} = -X^2(t)$$

Then

$$\frac{\delta X(t_1)}{\delta a(t)} = -X^2(t_1)$$

and

$$\frac{\delta R}{\delta a(t)} = -R^2 \quad (27)$$

in accordance with Eq. 20 obtained above using direct linearization.

Adjoint approach

In this approach, the operator L and right-hand term $\delta S - \delta N[X]$ of the linearized forward problem Eq. 21, along with the observables weighting function W of the linearized instrument model Eq. 23, are used in an alternative way involving the operator L^* , which is adjoint to L . This approach is more efficient when the number of parameters to be retrieved (counted by separate gridpoint

values for continuous parameters) exceeds the number of observables.

Rewriting for brevity the linearized forward model Eq. 21 as

$$LX' = S' \quad (28)$$

where $S' = \delta S - \delta N[X]$ and rewriting the linearized instrument model Eq. 23 as

$$R' = (W, X') \quad (29)$$

this approach can be outlined as follows. The adjoint operator L^* is derived from the requirement that the identity

$$(g, Lf) = (L^*g, f) \quad (30)$$

be satisfied for an arbitrary pair of functions f and g in the domain of L . Then, the solution X^* of the *adjoint problem*

$$L^*X^* = W \quad (31)$$

provides an alternative way to compute the linearized observables:

$$R' = (X^*, S') \quad (32)$$

This can be demonstrated by multiplying the adjoint problem by X' and the linearized forward problem by X^* and comparing the left and right sides of the resulting equalities:

$$(L^*X^*, X') = (W, X'), (X^*, LX') = (X^*, S')$$

Since the left-hand terms (L^*X^*, X') and (X^*, LX') are equal by definition of the adjoint operator, the right-hand terms (W, X') and (X^*, S') are also equal. Since $(W, X') = R'$, then $R' = (X^*, S')$.

Now, returning back from abbreviated Eq. 28 to Eq. 21 and replacing in Eq. 32 $R' \rightarrow \delta R$ and $S' \rightarrow \delta S - \delta N[X]$, we obtain a direct expression of the variation of the observables through variations of the operator and right-hand term of the forward problem:

$$\delta R = (X^*, \delta S - \delta N[X]) \quad (33)$$

This yields expressions for sensitivities of observables to discrete and continuous parameters in the form

$$\begin{aligned} \frac{\delta R}{\partial p_j} &= \left(X^*, \frac{\partial S}{\partial p_j} - \frac{\partial N}{\partial p_j} X \right), \text{ and } \frac{\delta R}{\delta p(x')} \\ &= \left(X^*, \frac{\delta S}{\delta p(x')} - \frac{\delta N}{\delta p(x')} X \right) \end{aligned} \quad (34)$$

As an example, consider a matrix forward problem

$$\begin{cases} \frac{d\mathbf{X}}{dt} + \mathbf{a}(t)\mathbf{X}(t) = \mathbf{b}(t) \\ \mathbf{X}(t_0) = \mathbf{X}_0 \end{cases} \quad (35)$$

where $\mathbf{a}(t)$ is a $n \times n$ -matrix and $\mathbf{X}(t)$, $\mathbf{b}(t)$, and $\mathbf{X}_0$ are n -vectors. The forward operator has the form

$$\mathbf{L}\mathbf{X} = \left(\frac{d}{dt} + \mathbf{a}(t) \right) \mathbf{X}(t) + \delta(t - t_0) \mathbf{X}(t) \quad (36)$$

To derive the operator adjoint to $\mathbf{L}$, we write down the left side of the identity Eq. 30 and transform it to the right side of this identity. For the left side of Eq. 30, we have

$$(\mathbf{X}^*, \mathbf{L}\mathbf{X}) = \int_{t_0}^{t_1} \mathbf{X}^{*T}(t) \left(\frac{d\mathbf{X}}{dt} + \mathbf{a}(t)\mathbf{X}(t) + \delta(t - t_0)\mathbf{X}(t) \right) dt$$

Performing integration by parts and canceling terms with opposite signs

$$\begin{aligned} & \int_{t_0}^{t_1} \mathbf{X}^{*T}(t) \frac{d\mathbf{X}}{dt} dt + \int_{t_0}^{t_1} \delta(t - t_0) \mathbf{X}(t) \mathbf{X}^{*T}(t) dt = \\ & - \int_{t_0}^{t_1} \frac{d\mathbf{X}^{*T}}{dt} \mathbf{X}(t) dt + \int_{t_0}^{t_1} \delta(t_1 - t) \mathbf{X}^{*T}(t) \mathbf{X}(t) dt \end{aligned}$$

we rewrite $(\mathbf{X}^*, \mathbf{L}\mathbf{X})$ in the form of the right side of Eq. 30:

$$\begin{aligned} & \int_{t_0}^{t_1} \left(-\frac{d\mathbf{X}^*}{dt} + \mathbf{a}^T(t)\mathbf{X}^*(t) + \delta(t_1 - t)\mathbf{X}^*(t) \right)^T \mathbf{X}(t) dt \\ & = (\mathbf{L}^* \mathbf{X}^*, \mathbf{X}) \end{aligned}$$

This yields the expression for the operator $\mathbf{L}^*$ adjoint to operator $\mathbf{L}^*$, Eq. 36:

$$\mathbf{L}^* \mathbf{X}^* = \left(-\frac{d}{dt} + \mathbf{a}^T(t) \right) \mathbf{X}^*(t) + \delta(t_1 - t) \mathbf{X}^*(t) \quad (37)$$

To formulate the adjoint problem in the form of Eq. 31, we assume observables as an m -vector $\mathbf{R}$. This yields the linearized instrument model in the form

$$\mathbf{R} = (\mathbf{W}, \mathbf{X}) = \int_{t_0}^{t_1} \mathbf{W}^T(t) \mathbf{X}(t) dt \quad (38)$$

with observables weighting function $\mathbf{W}(t)$ being an $m \times n$ -matrix. Assuming the vector $\mathbf{R} = \mathbf{W}_0 \mathbf{X}(t_1)$ is to be computed using a constant matrix $\mathbf{W}_0$ applied to the forward solution $X(t)$ at $t = t_1$, we have $\mathbf{W}(t) = \delta(t_1 - t) \mathbf{W}_0$. Corresponding adjoint problem has the form

$$\begin{cases} -\frac{d\mathbf{X}^*}{dt} + \mathbf{a}^T(t)\mathbf{X}^*(t) = 0 \\ \mathbf{X}^*(t_1) = \mathbf{W}_0^T \end{cases} \quad (39)$$

As an illustration, consider the nonlinear demo forward model Eqs. 2 and 7. Repeating above derivations for the operator of corresponding linearized forward problem, Eq. 25, we obtain

$$L^* X^* = \left(-\frac{d}{dt} + 2a(t)X^*(t) + \delta(t - t_1) \right) X^*(t) \quad (40)$$

Thus, corresponding adjoint problem has the form

$$\begin{cases} -\frac{dX^*}{dt} + 2a(t)X(t)X^*(t) = 0 \\ X^*(t_1) = 1 \end{cases} \quad (41)$$

Its solution (note that integration proceeds backward):

$$X^*(t) = \exp \left(-2 \int_t^{t_1} a(t') X(t') dt' \right)$$

To obtain sensitivities to parameters, we will use Eq. 33. For the sensitivity to the discrete parameter X_0 , we have

$$\begin{aligned} \delta_{X_0} R &= (X^*, \delta_{X_0}(S)) = \int_{t_0}^{t_1} X^*(t) \delta(t - t_0) \delta X_0 dt \\ &= X^*(t_0) \delta X_0 = \exp \left(-2 \int_{t_0}^{t_1} a(t') X(t') dt' \right) \delta X_0 \\ &= \left(\frac{R}{X_0} \right)^2 \delta X_0 \end{aligned}$$

Thus

$$\frac{\partial R}{\partial X_0} = \left(\frac{R}{X_0} \right)^2 \quad (42)$$

in accordance with Eqs. 19 and 26 obtained above using direct linearization and the linearization approach.

For the sensitivity to the continuous parameter $a(t)$, we have

$$\delta_a R = (X^*, -(\delta_a N)[X])$$

Rewriting $X^*(t)$

$$\begin{aligned} X^*(t) &= \exp \left(-2 \int_t^{t_1} a(t') X(t') dt' \right) \\ &= \exp \left[-2 \left(\int_{t_0}^{t_1} a(t') X(t') dt' - \int_{t_0}^t a(t') X(t') dt' \right) \right] \\ &= \frac{\exp \left(-2 \int_{t_0}^{t_1} a(t') X(t') dt' \right)}{\exp \left(-2 \int_{t_0}^t a(t') X(t') dt' \right)} = \frac{(X(t_1)/X_0)^2}{(X(t)/X_0)^2} = \frac{R^2}{X^2(t)} \end{aligned}$$

we have

$$\begin{aligned}\delta_a R &= (X^*, -(\delta_a N)[X]) = - \int_{t_0}^{t_1} X^*(t) X^2(t) \delta_p a(t) dt \\ &= -R^2 \int_{t_0}^{t_1} \delta a(t) dt\end{aligned}$$

Thus

$$\frac{\delta R}{\delta a(t)} = -R^2 \quad (43)$$

in accordance with Eqs. 20 and 27 obtained above using direct linearization and the linearization approach.

Summary

Forward modeling is a necessary component of geophysical retrieval algorithms. It includes the forward problem (the quantitative description of the subject of the geophysical retrieval itself) and the instrument model (the quantitative description of the procedure of measurements). In the most general case, the forward problem is represented by a system of differential equations with initial and/or boundary conditions, which, in general, can be solved only numerically. On the other hand, the instrument model is represented by an analytic expression, which converts the solution of the forward problem into observables, which simulate measured quantities used in geophysical retrievals.

Together, the forward model serves two functions. Besides simulated observables, it provides sensitivities of these observables with respect to the model parameters intended to be retrieved. This can be done in three different ways depending on the preferences of the researcher and on the relation between numbers of observables and model parameters to be retrieved. The finite-difference approach is most straightforward, but also least computer efficient at the same time. The linearization approach is more sophisticated, and it is preferable when the number of parameters to be retrieved is less than the number of observables. The adjoint approach is the most sophisticated and is preferable when parameters to be retrieved outnumber the observables, like in the retrievals of atmospheric profiles in atmospheric remote sensing.

A substantial amount of literature exists on forward modeling and sensitivity analysis (although under various names) in various areas of remote sensing, and the reader is encouraged, once he/she gets a big picture, to do independent search for this literature and to surf the Internet for this purpose. The references below will help in implementation of practical algorithms.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Korn, G. A., and Korn, T. M., 2000. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. New York: Dover Civil and Mechanical Engineering.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 2007. *Numerical Recipes*, 3rd edn. New York/Melbourne/Madrid/Cape Town/Singapore/São Paulo: Cambridge University Press. The Art of Scientific Computing.

GEOPHYSICAL RETRIEVAL, INVERSE PROBLEMS IN REMOTE SENSING

Eugene Ustinov

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition

Forward model. Quantitative tool for simulation of observables for a given set of model parameters.

Forward problem. System of differential equations with initial and/or boundary conditions, the solution of which, the *forward solution*, is used for simulation of observables.

Forward solution. Solution of the *forward problem*.

Instrument model. Quantitative procedure for evaluation of observables from the *forward solution*, which models the characteristics of the instrument and the procedure for measurements.

Inverse problem. Mathematical problem, usually in the form of a matrix equation, the solution of which, the *inverse solution*, represents an estimate of the *state vector*.

Model parameters. Parameters of the forward problem contained in its differential equations and initial and/or boundary conditions.

Observables. Output parameters of the *forward model*, which simulate the quantities measured in the geophysical retrieval.

Sensitivities. Partial or variational derivatives of *observables* with respect to *model parameters*.

State vector. Subset of the set of model parameters, which is intended to be retrieved.

Introduction

In contrast to *forward modeling*, which proceeds from assumed values of model parameters of the subject of study to simulated observables, the input in inverse modeling consists of measured observables, and the output consists of retrieved model parameters. Also, in contrast to forward modeling, where modeling of observables includes two

major components – *forward problem* (conversion of model parameters into the *forward solution*) and *instrument model* (conversion of the forward solution into simulated observables) – inverse modeling consists of only one major component, *inverse problem*, which converts the measured observables into estimates of the *state vector*, a subset of model parameters subject to the retrieval.

The value of any physical measurement – direct, like in situ measurements, or indirect, like in geophysical retrievals – has to be associated with its uncertainty, measurement error. Thus, it is highly desirable to retrieve the values of model parameters along with their *retrieval errors* (aka error bars). The existing methods of solution of the inverse problem deliver the solution along with its retrieval errors derived from errors of measurements driven by the instrument design.

Formulation and information analysis of the inverse problem

As stated in the accompanying article “*Forward Modeling*,” the quantitative relation between model parameters $\mathbf{p}$, and simulated observables $\mathbf{R}$ is, in general, nonlinear. The inverse problem is formulated, essentially, as inversion of this relation. Using high-level notations, we have:

$$\mathbf{F}(\mathbf{p}) = \mathbf{R} \quad (1)$$

It is assumed that all continuous model parameters are represented by their values on suitable grids of arguments and, correspondingly, $\mathbf{F}$ is a vector function of a vector argument.

The necessary premise for the formulation of the inverse problem is the availability of a computational procedure, which, for a given set of n model parameters described by a vector $\mathbf{p}$, provides a capability to compute a set of m observables described by a vector $\mathbf{R}$ along with their partial derivatives $\partial\mathbf{R}/\partial\mathbf{p}$ treated as a single $m \times n$ matrix of sensitivities $\mathbf{K}$. Another premise, which is necessary for the error analysis of the resulting inverse problem, as well as for its solution afterward, is knowledge of measurement errors – uncertainties of measured observables described by an m – vector $\boldsymbol{\epsilon}$. Computation of the corresponding covariance matrix is a necessary attribute of the algorithm of solution of the inverse problems in geophysical retrievals.

Differences between measured and simulated observables are commonly referred to as *residuals*. The nonlinear inverse problem is solved iteratively, and at each iteration, the solution of the corresponding linearized inverse problem yields corrections to the current approximation of the set of model parameters. These corrections are commonly referred to as *increments*.

The nonlinear inverse problem is solved iteratively. The corresponding linearized inverse problem in high-level notations has the form:

$$\mathbf{K}\mathbf{x} = \mathbf{y} \quad (2)$$

where $\mathbf{y} = \mathbf{R}_{obs} - \mathbf{R}$ is the m – vector of residuals and $\mathbf{x} = \mathbf{p}_{next} - \mathbf{p}$ is the n – vector of increments.

Based on the sensitivity matrix $\mathbf{K}$, a preliminary information analysis of the inverse problem can be conducted. In general, not all elements of the measurement set are independent from each other, and, as a result, the number of linearly independent rows of $\mathbf{K}$ may be less than the total number of the rows. The number $r \leq m$ of linearly independent rows is referred to as the *rank* of the matrix $\mathbf{K}$.

Rank of matrix $\mathbf{K}$ can be determined using a technique called singular value decomposition. An eigenvalue problem is considered in the form:

$$\begin{pmatrix} \mathbf{0} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \text{ or, in explicit form :} \\ \mathbf{K}\mathbf{v} = \lambda\mathbf{u} \\ \mathbf{K}^T\mathbf{u} = \lambda\mathbf{v} \quad (3)$$

Here $\mathbf{u}$ is an m – vector in the space of measurement vectors and $\mathbf{v}$ is an n – vector in the space of state vectors. This problem can be rewritten in the form of two eigenvalue problems for matrices $\mathbf{K}^T\mathbf{K}(n \times n)$ and $\mathbf{K}\mathbf{K}^T(m \times m)$:

$$\mathbf{K}^T\mathbf{K}\mathbf{v} = \lambda^2\mathbf{v} \quad (4)$$

$$\mathbf{K}\mathbf{K}^T\mathbf{u} = \lambda^2\mathbf{u} \quad (5)$$

which are solved by standard methods. The numbers of resulting nonzero eigenvalues of the matrices $\mathbf{K}^T\mathbf{K}$ and $\mathbf{K}\mathbf{K}^T$ coincide and yield the rank p of the matrix $\mathbf{K}$.

Number of degrees of freedom (DOF) for signal d_s is another quantitative measure of quality of the measurements selected for the retrieval. It indicates how many independent quantities can be measured. In general, it is not an integer number:

$$d_s = \text{tr} \left(\Lambda^2 (\Lambda^2 + \mathbf{I}_m)^{-1} \right) = \sum_{i=1}^m \frac{\lambda_i^2}{\lambda_i^2 + 1} \quad (6)$$

An indicator complementary to d_s is the number of degrees of freedom for noise d_n :

$$d_n = \text{tr} \left((\Lambda^2 + \mathbf{I}_m)^{-1} \right) = \sum_{i=1}^m \frac{1}{\lambda_i^2 + 1} \quad (7)$$

Shannon information content H of measurements used in retrievals with a given matrix $\mathbf{K}$ can be estimated assuming Gaussian noise:

$$H = \frac{1}{2} \sum_{i=1}^m \ln(1 + \lambda_i^2) \quad (8)$$

Solution of Inverse Problems

Unlike solution of forward problems, which, in principle, can be accomplished with any numerical accuracy, solution of inverse problems is associated with inherent uncertainty, for two general reasons: presence of measurement errors, and the indirect, inference-like nature of the process of retrieval, which is implemented by solution of the inverse problem.

The presence of measurement errors forces to consider the residuals vector $\mathbf{y}$ as a random quantity. It is customary to replace $\mathbf{y} \rightarrow \mathbf{y} + \boldsymbol{\varepsilon}$, where the vector of measurement errors $\boldsymbol{\varepsilon}$ is assumed to obey the Gaussian distribution with an average (mathematical expectation) $\bar{\boldsymbol{\varepsilon}} = 0$ and a covariance matrix $\mathbf{C}_y$. The corresponding *probability distribution function* (PDF) has the form:

$$P(\boldsymbol{\varepsilon}) \propto \exp\left(-\frac{1}{2} \boldsymbol{\varepsilon}^T \mathbf{C}_y^{-1} \boldsymbol{\varepsilon}\right) \quad (9)$$

where the symbol “ $\propto$ ” means proportionality. In the case of non-correlating measurement errors $s_i (i = 1, \dots, m)$ Eq. 9 can be rewritten as

$$P(\boldsymbol{\varepsilon}) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^m \frac{\varepsilon_i^2}{s_i^2}\right) \quad (10)$$

With measurement errors included explicitly, the inverse problem Eq. 2 takes the form

$$\mathbf{K}\mathbf{x} = \mathbf{y} + \boldsymbol{\varepsilon} \quad (11)$$

where the vector $\mathbf{x}$ is treated as a random quantity, which also obeys the Gaussian distribution with the PDF:

$$P(\mathbf{x}) \propto \exp\left(-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{C}_x^{-1} (\mathbf{x} - \bar{\mathbf{x}})\right) \quad (12)$$

The solution of the inverse problem Eq. 2 is sought as the average $\bar{\mathbf{x}}$ of this distribution, whereas the covariance matrix $\mathbf{C}_x$ describes uncertainty (retrieval errors) of this solution.

By definition of the covariance matrix, $\mathbf{C}_x$ is symmetric, i.e., $\mathbf{C}_x = \mathbf{C}_x^T$, and $\mathbf{C}_x^{-1} = (\mathbf{C}_x^{-1})^T$. Then, the PDF Eq. 12 can be transformed as

$$P(\mathbf{x}) \propto \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{B} \mathbf{x} + \mathbf{b}^T \mathbf{x}\right) \quad (13)$$

where

$$\mathbf{B} = \mathbf{C}_x^{-1} \quad (14)$$

and

$$\mathbf{b} = \mathbf{B} \bar{\mathbf{x}} \quad (15)$$

On the other hand, the PDF $P(\mathbf{x})$ can be derived from the PDF $P(\boldsymbol{\varepsilon})$ Eq. 9 by the substitution $\boldsymbol{\varepsilon} = \mathbf{K}\mathbf{x} - \mathbf{y}$. Essentially, this is an *a posteriori* PDF for the random quantity $\mathbf{x}$ with quantity $\mathbf{y}$ known:

$$P(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2} (\mathbf{K}\mathbf{x} - \mathbf{y})^T \mathbf{C}_y^{-1} (\mathbf{K}\mathbf{x} - \mathbf{y})\right) \quad (16)$$

We have:

$$\begin{aligned} & (\mathbf{K}\mathbf{x} - \mathbf{y})^T \mathbf{C}_y^{-1} (\mathbf{K}\mathbf{x} - \mathbf{y}) \\ &= (\mathbf{K}\mathbf{x})^T \mathbf{C}_y^{-1} \mathbf{K}\mathbf{x} - (\mathbf{K}\mathbf{x})^T \mathbf{C}_y^{-1} \mathbf{y} - \mathbf{y}^T \mathbf{C}_y^{-1} \mathbf{K}\mathbf{x} + \mathbf{y}^T \mathbf{y} \end{aligned} \quad (17)$$

Note that all terms in Eq. 17 are scalars, and thus are equal to themselves transposed. In particular

$$\mathbf{y}^T \mathbf{C}_y^{-1} \mathbf{K}\mathbf{x} = (\mathbf{K}\mathbf{x})^T \mathbf{C}_y^{-1} \mathbf{y}$$

Observing also that the term $\mathbf{y}^T \mathbf{y}$ in Eq. 17 does not depend on $\mathbf{x}$ we can rewrite Eq. 16 as:

$$P(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} \mathbf{x} + \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{y}\right) \quad (18)$$

Comparing Eq. 18 with the general form of PDF Eq. 13, we obtain:

$$\mathbf{B} = \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} \quad (19)$$

and

$$\mathbf{b} = \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{y} \quad (20)$$

Comparison with the equality Eq. 15 yields a matrix equation for the solution $\bar{\mathbf{x}}$ of the inverse problem Eq. 2:

$$\mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} \bar{\mathbf{x}} = \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{y} \quad (21)$$

along with the expression for the covariance matrix of this solution:

$$\mathbf{C}_x = \left(\mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K}\right)^{-1} \quad (22)$$

In the particular case of non-correlated equal measurement errors, when $s_i \equiv s$, the matrix equation Eq. 21 and expression for the covariance matrix of its solution Eq. 22 reduce to:

$$\mathbf{K}^T \mathbf{K} \bar{\mathbf{x}} = \mathbf{K}^T \mathbf{y} \quad (23)$$

and

$$\mathbf{C}_x = s^2 (\mathbf{K}^T \mathbf{K})^{-1} \quad (24)$$

If the PDF of $\mathbf{x}$ can be sufficiently constrained based on some additional, *a priori* information, then the number of

measurements m can be less than the dimension n of $\mathbf{x}$. This is the case, e.g., in atmospheric remote sensing, when information about correlation between values of the atmospheric parameter to be retrieved can be invoked in the form of a covariance matrix

$$\mathbf{C}_{\mathbf{x},a} = \overline{\mathbf{x}\mathbf{x}^T} \quad (25)$$

Then the a priori PDF of $\mathbf{x}$ can be represented in the form of the corresponding Gaussian distribution

$$P_a(\mathbf{x}) \propto \exp\left(-\frac{1}{2} \mathbf{x}^T \mathbf{C}_{\mathbf{x},a}^{-1} \mathbf{x}\right) \quad (26)$$

and the resulting constrained PDF is a product of PDFs Eqs. 18 and 26:

$$P_c(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}|\mathbf{y}) \cdot P_a(\mathbf{x}) \\ \propto \exp\left(-\frac{1}{2} \mathbf{x}^T (\mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1}) \mathbf{x} + \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{y}\right) \quad (27)$$

Correspondingly, the matrix equation for the solution $\bar{\mathbf{x}}$ of the *regularized* inverse problem Eq. 2 and expression for the covariance matrix $\mathbf{C}_x$ of this solution, respectively, take the form:

$$(\mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1}) \bar{\mathbf{x}} = \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{y} \quad (28)$$

$$\mathbf{C}_x = (\mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1})^{-1} \quad (29)$$

In practice, the number of measurements m in atmospheric remote sensing can be larger, and even much larger than n , e.g., in retrievals from data of high-resolution spectral measurements with the Tropospheric Emissions Spectrometer (TES) flown on the Aura spacecraft. But the rank r of the corresponding sensitivity matrix $\mathbf{K}(m \times n)$ is of the order of the number of atmospheric scale heights over the altitude range covered by retrievals. Of course, the number of necessary grid points n is substantially larger, and, thus, invoking of a priori information is necessary.

Recall that in general the inverse problem $\mathbf{F}(\mathbf{p}) = \mathbf{R}$ is nonlinear and it has to be solved using the linearized inverse problem $\mathbf{K}\mathbf{x} = \mathbf{y}$. If the first guess of the state vector $\mathbf{p}_0$ is too far from the solution, then a straightforward application of the above approach at each iteration may not converge to the solution. A more robust approach named Levenberg–Marquardt method, which has a long history of development (see Rodgers (2000) for a review), makes it possible to cope with this difficulty. At each iteration, the size of the step $\bar{\mathbf{x}}$ is regulated by introducing an additional matrix term proportional to a scalar γ , which is chosen based on some semiempirical rules. If no regularization is necessary, this matrix term is $\gamma\mathbf{D}$, where $\mathbf{D}$ is a diagonal scaling matrix with elements corresponding to magnitudes and dimensions of the elements of the state vector $\mathbf{p}$. If the vector $\mathbf{p}$ consists of elements of same magnitudes and dimensions, then $\mathbf{D}$ reduces to the identity matrix $\mathbf{I}$. If regularization is

necessary, Rodgers (2000) suggests this term in the form $\gamma\mathbf{C}_{\mathbf{x},a}^{-1}$. Thus, the step $\bar{\mathbf{x}}$ is sought as a solution of the above matrix equations modified accordingly:

$$(\mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} + \gamma\mathbf{D}) \bar{\mathbf{x}} = \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{y} \quad (30)$$

or

$$(\mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} + (1 + \gamma)\mathbf{C}_{\mathbf{x},a}^{-1}) \bar{\mathbf{x}} = \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{y} \quad (31)$$

Error analysis of inverse problems

The covariance matrix $\mathbf{C}_x$ provides a straightforward way for estimations of the retrieval errors (error bars) of retrieved profiles defined as variances derived from the covariance matrix

$$\sigma_j = \sqrt{(x_j - \bar{x}_j)^2} = \sqrt{(\mathbf{C}_x)_{jj}}, \quad (j = 1, \dots, n) \quad (32)$$

When additional information about the solution in the form of the a priori covariance matrix $\mathbf{C}_{\mathbf{x},a}$ is used, then additional, *smoothing* errors emerge. Whereas error bars represent uncertainties of retrieved values, the smoothing errors represent uncertainties of benchmarking of retrieved values, e.g., retrieved atmospheric profiles, to specific reference arguments, e.g., altitude in the atmosphere. This smoothing error is represented by the $n \times n$ matrix $\mathbf{A} = \partial\bar{\mathbf{x}}/\partial\mathbf{x}$ called the *averaging kernel*. Using the substitution $\mathbf{y} = \mathbf{K}\mathbf{x}$ in the right-hand term of the regularized inverse problem in Eq. 28, and solving it for $\bar{\mathbf{x}}$, we have:

$$\bar{\mathbf{x}} = (\mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1})^{-1} \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} \mathbf{x}$$

Thus the averaging kernel $\mathbf{A}$ has the form:

$$\mathbf{A} = (\mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1})^{-1} \mathbf{K}^T \mathbf{C}_y^{-1} \mathbf{K} \quad (33)$$

It should be emphasized that the accuracy of estimation of the smoothing error represented by the matrix $\mathbf{A}$ depends on the accuracy of the knowledge of a priori covariance matrix $\mathbf{C}_{\mathbf{x},a}$. On the other hand, if the solution does not need to be constrained, then a priori PDF $P(\mathbf{x}) \propto 1$ and correspondingly $\mathbf{C}_{\mathbf{x},a}^{-1} = 0$, and the averaging kernel $\mathbf{A}$ reduces to the $n \times n$ identity matrix $\mathbf{I}_n$.

Finally, retrieval errors may be associated with uncertainties in parameters of the forward model, which are not being retrieved and have to be assumed based on some independent information. General analysis of associated uncertainties is given in the monograph of Clive Rodgers (2000).

Summary

Formulation of the inverse problem, analysis of its information content, choice of the method of its solution, and analysis of resulting retrieval errors represent main phases of development of the inversion algorithm in

geophysical retrievals. The key premise is availability of the adequate forward model, which, for a given set of model parameters, provides accurate values of simulated observables along with the matrix of sensitivities of observables with respect to elements of the state vector. This makes it possible to formulate the linearized inverse problem. If this problem is ill posed, then the algorithm of its solution involves regularization using some a priori information about this solution. The solution is sought as an average over a probability distribution, which is driven by the probability distribution of measurement errors around the measured observables and by the matrix of sensitivities provided by the forward model. The variance of the probability distribution of the solution provides the estimate of retrieval errors. If necessary, this estimate needs to be complemented by estimates of errors due to uncertainties of the model parameters outside the state vector and due to the approximate nature of the forward model.

There is a vast amount of literature describing formulation and solution of inverse problems in geophysical retrievals. The reader is encouraged, once he/she gets a big picture, to do an independent search for this literature and to surf the Internet for this purpose. The monograph by Clive Rodgers (2000) is a rich source of information on formulation and solution of inverse problems in atmospheric remote sensing, which, as a rule, are ill-posed and need regularization. Two other monographs provide valuable information on various aspects of practical implementation of algorithms of solution of inverse problems.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Korn, G. A., and Korn, T. M., 2000. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. New York: Dover Civil and Mechanical Engineering.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 2007. *Numerical Recipes*, 3rd edn. New York/Melbourne/Madrid/Cape Town/Singapore/São Paulo: Cambridge University Press/The Art of Scientific Computing.
- Rodgers, C. D., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Singapore: World Scientific.

GEOPHYSICAL RETRIEVAL, OVERVIEW

Eugene Ustinov
Jet Propulsion Laboratory, California Institute of
Technology, Pasadena, CA, USA

Definition

Forward model. Quantitative tool for simulation of *observables* for a given set of *model parameters*.

Forward problem. System of differential equations with initial and/or boundary conditions, solution of which, the *forward solution*, is used for simulation of *observables*.

Forward solution. Solution of the *forward problem*.

Instrument model. Quantitative procedure for evaluation of observables from the *forward solution*, which models the characteristics of the instrument and the procedure for measurements.

Inverse problem. Mathematical problem, usually in the form of a matrix equation, solution of which, the *inverse solution*, represents an estimate of the *state vector*.

Model parameters. Parameters of the forward problem contained in its differential equations and initial and/or boundary conditions.

Observables. Output parameters of the forward model, which simulate the quantities measured in the geophysical retrieval.

Sensitivities. Partial or variational derivatives of *observables* with respect to *model parameters*.

State vector. Subset of the set of model parameters, which is intended to be retrieved.

Introduction

Geophysical retrievals are indirect measurements. In contrast to direct measurements, where there is a direct relation between the measured quantity (e.g., length of the mercury column in the thermometer and inferred quantity – temperature), in geophysical retrievals, the quantities to infer (retrieve) are related to measured quantities via an intermediary, a *forward model*, which provides a cause-to-effect relation between quantities to infer (a subset of *model parameters* constituting the *state vector*) and measured quantities (*observables*). For example, in atmospheric remote sensing, the parameters of the atmosphere/surface system are retrieved using measurements of radiances measured at the top of the atmosphere. Thus, in the retrieval process, we proceed back, from the effect, i.e., observed radiances, to the cause – atmospheric and surface parameters – specific values of which result in specific values of the measured radiances, which are used for the retrieval.

To adequately describe the dependence of measured quantities on the model parameters, the forward model has to have two mandatory components. Its first component is the *forward problem* – the description of the subject of retrieval, based on known physics of it, usually, in the form of a system of differential equations with corresponding initial and/or boundary conditions. In the case considered above, the forward problem consists of the equation of radiative transfer (RT) with boundary conditions complemented by known dependencies of radiative parameters directly entering the RT equation and boundary conditions, such as the Planck function, on geophysical parameters of interest, such as atmospheric temperature. In high-level notation, the forward problem can be written in the form of an operator equation:

$$LX = S$$

where the operator L denotes all homogeneous operations on the solution X contained in the differential equations and initial and/or boundary conditions of the forward problem and the right-hand term S combines all nonhomogeneous right-hand terms of the equations and conditions.

The second component of the forward model is the *instrument model*. In the above case, this model consists of angular and spectral distributions of the instrument's response. Convolution of the solution of the forward problem with functions describing the instrument model provides observables, which simulate measured quantities. In high-level notation, the expression for observables can be written in the form of an inner product of the forward solution X with the function W describing the response of the instrument:

$$R = (W, X)$$

Besides observables, the forward model provides another entity – *sensitivities* of observables to the parameters to be retrieved. Essentially, sensitivities are derivatives of observables with respect to the model parameters to be retrieved. For example, in atmospheric remote sensing, temperature weighting functions are sensitivities of observed radiances with respect to temperature. Sensitivities computed along with observables provide premises necessary to solve the *inverse problem*. In contrast to the forward model, the input of the inverse problem is the set of measured observables, and its output is the set of retrieved model parameters providing the best fit of computed observables to measured observables. Solution of the inverse problem implements the aforementioned path from effect to cause. This circumstance is the key reason of complications associated with the solution of the inverse problems.

Forward model: forward problem and sensitivity analysis

Solution of the forward problem and computation of sensitivities constitute two major components of the algorithm implementing the forward model. There are many software packages available, which make implementation of the forward problem in the forward model to be rather a craft than a skill. On the other hand, computation of sensitivities can be implemented in a number of ways, depending on preferences of the algorithm developer.

The simplest but most computer-intensive approach to computation of sensitivities is termed the *finite-difference approach*. The forward model is run once to obtain the baseline solution for the given set of values of model parameters. Then, the parameters to be retrieved are varied one by one, gridpoint by gridpoint. For each such variation, the forward problem is run separately. Resulting variations of observables are divided by variations of parameters yielding the sensitivity matrix – a matrix of partial derivatives of observables with respect to the model parameters to be retrieved.

Two other approaches require some mathematical skills and additional computer programming. For their application, the forward problem and instrument model of the forward model need to be linearized around the baseline solution. The resulting *linearized forward problem* and *linearized instrument model* can be used in two ways, complementing each other, depending on the relation between the number of observables and the number of retrieval parameters.

The *linearization approach* is based on the linearized forward problem formulated for derivatives of the baseline solution with respect to retrieval parameters. The linearized instrument model is applied to obtained derivatives of the baseline solution to calculate partial derivatives of observables with respect to retrieval parameters. The key point is that only right-hand terms of equations (differential equations and initial and/or boundary conditions) in the linearized forward problem vary with the specific retrieval parameter. The left-hand operations on the derivatives of the solution are the same. This provides substantial savings as compared to the finite-difference approach. An illustrative analogy here is the solution of a system of linear algebraic equations for a number of cases where only right-hand inhomogeneous terms vary, but coefficients in the left-hand homogeneous terms remain the same.

The *adjoint approach* is also based on the use of the linearized forward problem and linearized instrument model, but in a different way. It requires more mathematical skills than the linearization approach, as well as additional computer programming. All left-hand operations in the linearized forward problem are combined into a linear operator of this problem. The corresponding adjoint operator is derived based on the definition of this operator. The adjoint problem is formulated using the obtained adjoint operator with the linearized instrument models for given observables as right-hand terms. It turns out that the corresponding *adjoint solution* provides a way to obtain sensitivities with respect to all model parameters at once. Similarly to the linearization approach, all left-hand operations in the adjoint problem are the same, and only right-hand terms vary with the specific observable.

Comparison of performance of the linearization and adjoint approaches is based, essentially, on the comparison of the number of observables with the number of retrievable parameters. It is important to point out here that in the case of continuous parameters, such as atmospheric parameters, the value at each gridpoint should be counted as a separate parameter. With this in mind, if the number of observables exceeds the number of retrievable parameters, then the linearization approach is more efficient. In the opposite case, the adjoint approach is more efficient. For example, in many cases of atmospheric remote sensing, the observables are spectral radiances: a single radiance or some number of them observed at different nadir angles. The retrieval parameters are atmospheric profiles, usually specified on a few tens of gridpoints in the atmosphere. Objectively, the adjoint approach is more efficient

here. But historically, the linearization approach is still dominating, apparently because it is easier to comprehend and implement.

In a few practical situations, the forward problems can be solved analytically. For example, this is the case for radiative transfer of thermal radiation in the non-scattering atmosphere, which may be a good approximation, especially in the microwave spectral region. Then, observed radiances are also available analytically, and corresponding expressions can be linearized to yield sensitivities to desired parameters to be retrieved. The equation of radiative transfer is linear per se, so no linearization is necessary. Application of the adjoint approach in this case results in the adjoint problem, which also can be solved analytically. Of course, resulting analytic expressions for sensitivities are identical to those obtained directly.

Inverse problems and their solution

As was pointed out in the Introduction, the forward models proceed from model parameters (cause) to observables (effect) in two steps. The first step is implemented by the forward problem, which yields a forward solution providing a quantitative description of the subject of retrieval, per se, as is. This description is independent on the instrument model, which implements the second step of the forward model. Application of the instrument model to the forward solution yields simulated observables.

In the inverse process of retrieval, we proceed from observables to the subset of model parameters intended to be retrieved, usually referred to as the *state vector*. Proceeding from the effect to the cause does result in specific complications, usually nonexistent in the process of forward modeling. In general, inverse problems are nonlinear and have to be solved iteratively. At each step, a corresponding linearized inverse problem is solved, which has the form of a matrix equation

$$\mathbf{K}(\mathbf{p}_{k+1} - \mathbf{p}_k) = \mathbf{R}_{obs} - \mathbf{R}_k$$

where the vector of observables $\mathbf{R}_k$ and matrix of sensitivities $\mathbf{K}$ are calculated for the current iteration of the state vector $\mathbf{p}_k$, $\mathbf{R}_{obs}$ is the vector of measured observables, and $\mathbf{p}_{k+1}$ is the next iteration of the state vector. Depending on the physics of the forward model, there are two types of inverse problem: well-posed problems and ill-posed problems.

Well-posed problems do not require any additional information about the state vector, other than that provided by the measured observables themselves. The number of independent observables has to be no less than the number of model parameters in the state vector. Additional observables – dependent or independent – will result in decrease of retrieval errors.

Ill-posed problems cannot be solved in principle without invoking some additional information about the state vector known prior to retrieval, referred to as *a priori information*. This happens, for example, when the state

vector includes continuous parameters, which are functions of relevant arguments – usually, space coordinates and/or time. Vertical profiles of atmospheric parameters provide a good example. Another example, although not from the area of geophysics, is the trajectory of a celestial object – its position in the selected system of coordinates as a function of time.

The a priori information can be of various forms. In such case as retrievals of profiles of atmospheric parameters, this information can be provided in the form of covariance matrices of these parameters – derived from statistics of previous measurements, or just guessed from some considerations. In such case as trajectory retrievals, the a priori information can be in the form of requirements on smoothness of the solution in terms of the time derivative of a suitable order.

In most practical cases, the forward models are nonlinear and so are the corresponding inverse problems. They have to be linearized and solved iteratively. At each iteration, the solution sought is the correction of the state vector providing the best fit to the residuals – differences between measured and simulated observables. The linearized inverse problem has a form of a linear algebraic matrix equation, which can be solved using existing numerical recipes.

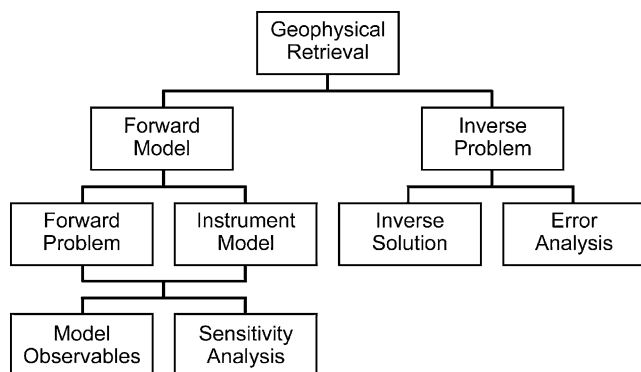
Any physical measurement is meaningful if and only if its result is obtained along with an estimate of its uncertainty. Thus, the solution of inverse problems in geophysical retrievals needs to be complemented by its error analysis. If the forward model accurately describes the subject of retrieval and the procedure of measurement and the resulting inverse problem is well-posed, then the measurement errors are the only source of retrieval errors. Resulting retrieval errors are commonly referred to as *error bars*.

If the inverse problem is ill-posed, then invoking a priori information results in an additional kind of retrieval errors. For example, vertical profiles retrieved in atmospheric remote sensing have finite vertical resolution, i.e., there are uncertainties in benchmarking of retrieved values against the vertical coordinate in the atmosphere. Resulting retrieval errors are referred to as *smoothing errors*.

Finally, any forward model represents only an approximation of the reality, and corresponding *forward model errors* contribute to retrieval errors. If formulation of the forward model itself is accurate, then the only source of forward model errors is due to uncertainties of model parameters outside of the state vector subset. But the forward model may contain uncertainties in its formulation. As Clive Rodgers states in his monograph (Rodgers, 2000), “then modeling error can be tricky to estimate.”

Summary

A necessary premise for development of geophysical retrieval algorithms is an availability of an adequate forward model, which, for the given set of model parameters, provides accurate values of simulated observables along with sensitivities of the observables to the state vector.



Global Climate Observing System, Figure 1 Block diagram of principal components of geophysical retrieval.

Another necessary premise is an availability of the adequately formulated inverse problem, solution of which provides the retrieval – a best estimate of the state vector corresponding to measured values of observables – along with estimates of uncertainties of this retrieval due to measurement errors, specifics of the inverse problem, and quality of the forward model. An assembly of the forward model and inverse problem into a robust framework of the algorithm of geophysical retrieval is a challenge, and success in meeting this challenge is a necessary premise of the success of any remote sensing project.

Principal components of geophysical retrieval discussed in this entry are illustrated in [Figure 1](#). (For further details, see [Geophysical Retrieval, Forward Models in Remote Sensing](#) and [Geophysical Retrieval, Inverse Problems in Remote Sensing](#)).

There is a vast amount of literature describing geophysical retrieval as a whole, as well as different aspects of it, both in forward modeling and in solution of corresponding inverse problems. A short list below is notably incomplete and reflects personal preferences of the author. The reader is encouraged, once he/she gets a big picture to do independent search for this literature and to surf the Internet for this purpose.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Korn, G. A., and Korn, T. M., 2000. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. New York: Dover Civil and Mechanical Engineering.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 2007. Numerical recipes. In *The Art of Scientific Computing*, 3rd edn. New York/Melbourne/Madrid/Cape Town/Singapore/Saõ Paulo: Cambridge University Press.
- Rodgers, C. D., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Singapore: World Scientific.

GLOBAL CLIMATE OBSERVING SYSTEM

Jean-Louis Fellous

Committee on Space Research (COSPAR) Secretariat,
Paris, France

Definition

Albedo. The fraction of solar energy reflected from the Earth back into space.

Fraction of absorbed photosynthetically active radiation (FAPAR). The solar radiation reaching the surface on the 0.4–0.7 μm spectral region is known as the photosynthetically active radiation (PAR). FAPAR refers to the fraction of PAR that is absorbed by a vegetation canopy.

Leaf area index. The ratio of total upper leaf surface of vegetation divided by the surface area of the land on which the vegetation grows.

Permafrost. Soil at or below the freezing point of water (0°C or 32°F) for 2 or more years.

Phytoplankton. Photosynthetic or plant constituent of plankton, mainly composed of unicellular algae.

Introduction

The Global Climate Observing System (GCOS) was established in 1992. GCOS is an international mechanism aimed at coordinating observing systems and networks for meeting the needs for climate observation at national and global level. GCOS serves as the climate component of the *Global Earth Observation System of Systems (GEOSS)*. It is cosponsored by four international bodies: the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission (IOC) of UNESCO, the United Nations Environment Programme (UNEP), and the International Council for Science (ICSU).

GCOS objectives

GCOS's main role is to ensure that the observations required to meeting national and international needs for climate data and information are well identified, gathered, and made widely available. Its ultimate goal is the establishment of a global, reliable, comprehensive, and sustained climate observing system, giving access to the physical, chemical, and biological properties as well as the atmospheric, oceanic, hydrological, terrestrial, and cryospheric processes that contribute the total Earth climate system and capable of detecting and monitoring its natural and man-induced changes.

GCOS activities provide the necessary support to the components of the World Climate Programme, including the World Climate Research Programme, to the assessment role of the Intergovernmental Panel on Climate Change (IPCC), and to the international policy development undertaken through the United Nations Framework Convention on Climate Change (UNFCCC).

GCOS builds on the existing observing systems operated in the context of the WMO World Weather Watch Global Observing System and Global Atmosphere Watch, the *Global Ocean Observing System (GOOS)* led by the IOC, and the *Global Terrestrial Observing System (GTOS)* led by the Food and Agriculture Organization (FAO) and associated Global Terrestrial Networks. It essentially relies upon observing system elements funded and operated at national level. It includes in situ, airborne, and space-based observational components.

GCOS implementation

A GCOS Steering Committee is charged with the responsibility of providing scientific and technical guidance to the sponsoring and participating organizations for the planning and implementation of the observing system. Scientific advisory panels have also been established to provide expert advice in each domain (atmosphere, ocean, and land surface). These panels are consulted on the appropriate observing strategy and the measurement requirements and contribute to assessing the status of observing networks and systems. A Secretariat located at the WMO Headquarters in Geneva, Switzerland, supports the activities of the Steering Committee, the GCOS panels and the GCOS program as a whole.

The overall plan for the global climate observing system was developed over the 1992–1995 period. GCOS, in consultation with its partners, further prepared and published in 2004 an Implementation Plan (GCOS, 2004) that addressed the requirements identified in the Second Report (GCOS, 2003) on the Adequacy of Global Observing Systems for Climate in Support of the United Nations Framework Convention on Climate Change. This plan specifically responded to the request of the Conference of the Parties to the UNFCCC. It was further complemented in 2006 with a Supplement (GCOS, 2006) specifically dedicated to space-based observations.

Essential climate variables

The GCOS Implementation Plan (GCOS-IP) is based on the needs and requirements relating to the observation of a key list of “essential climate variables” (see Table 1 below) that were identified as technically and economically feasible to observe and having critical impact with respect to the scientific requirements for systematic climate monitoring. The GCOS-IP specified 131 actions in the atmospheric, oceanic, and terrestrial domains to address the observing system needs of the UNFCCC. It outlined a comprehensive program to implement these actions by national, regional, and international entities and focused on needed improvements for observation of 44 essential climate variables (ECVs) for the atmosphere, oceans, and land.

Climate monitoring principles

Global climate change provides a specific challenge for climate monitoring, both through the need for global

Global Climate Observing System, Table 1 Essential climate variables

Domain	Essential climate variables	
Atmospheric (over land, sea, and ice)	Surface	Air temperature, precipitation, air pressure, surface radiation budget, wind speed and direction, water vapor
	Upper air	Earth radiation budget (including solar irradiance), upper-air temperature, wind speed and direction, water vapor, cloud properties
	Composition	Carbon dioxide, methane, ozone, other long-lived greenhouse gases, aerosol properties
Oceanic	Surface	Sea-surface temperature, sea-surface salinity, sea level, sea state, sea ice, current, ocean color (for biological activity), carbon dioxide partial pressure
	Subsurface	Temperature, salinity, current, nutrients, carbon, ocean tracers, <i>phytoplankton</i>
Terrestrial	River discharge, water use, ground water, lake levels, snow cover, glaciers and ice caps, permafrost and seasonally frozen ground, albedo, land cover (including vegetation type), fraction of absorbed photosynthetically active radiation, leaf area index, biomass, fire disturbance, soil moisture	

coverage and because the rate of change of climate variables (such as average temperature, sea level, and rainfall) tends to be small compared with the “background noise” of natural climate variability. Thus, particular attention to the quality and consistency of observations is needed for climate monitoring. In an attempt to help extend the compliance with the general standards and good practices for climate observation, GCOS has thus developed a set of ten “climate monitoring principles” for the collection, archival, and analysis of observations, which were endorsed by the UNFCCC in 1999. GCOS further proposed ten additional satellite monitoring principles that were agreed by the world’s space agency members of the Committee on Earth Observation Satellites (CEOS) in 2003. These principles are reproduced in Table 2 below.

The status of the global climate observing system

In December 2005, the GCOS Secretariat was invited to provide a comprehensive report in June 2009 on progress with actions recommended in the GCOS-IP to maintain, strengthen, or otherwise facilitate global observations of the climate system, including adherence to the GCOS Climate Monitoring Principles. In response to this invitation, a report was submitted to the UNFCCC in April 2009 and

Global Climate Observing System, Table 2 Climate monitoring principles

Effective monitoring systems for climate should adhere to the following principles:

1. The impact of new systems or changes to existing systems should be assessed prior to implementation
2. A suitable period of overlap for new and old observing systems is required
3. The details and history of local conditions, instruments, operating procedures, data processing algorithms, and other factors pertinent to interpreting data (i.e., metadata) should be documented and treated with the same care as the data themselves
4. The quality and homogeneity of data should be regularly assessed as a part of routine operations
5. Consideration of the needs for environmental and climate monitoring products and assessments, such as IPCC assessments, should be integrated into national, regional, and global observing priorities
6. Operation of historically uninterrupted stations and observing systems should be maintained
7. High priority for additional observations should be focused on data-poor regions, poorly observed parameters, regions sensitive to change, and key measurements with inadequate temporal resolution
8. Long-term requirements, including appropriate sampling frequencies, should be specified to network designers, operators, and instrument engineers at the outset of system design and implementation
9. The conversion of research observing systems to long-term operations in a carefully planned manner should be promoted
10. Data management systems that facilitate access, use, and interpretation of data and products should be included as essential elements of climate monitoring systems
Furthermore, operators of satellite systems for monitoring climate need to:
 - (a) Take steps to make radiance calibration, calibration monitoring, and satellite-to-satellite cross-calibration of the full operational constellation a part of the operational satellite system
 - (b) Take steps to sample the Earth system in such a way that climate-relevant (diurnal, seasonal, and long-term interannual) changes can be resolved. Thus satellite systems for climate monitoring should adhere to the following specific principles:
11. Constant sampling within the diurnal cycle (minimizing the effects of orbital decay and orbit drift) should be maintained
12. A suitable period of overlap for new and old satellite systems should be ensured for a period adequate to determine inter-satellite biases and maintain the homogeneity and consistency of time-series observations
13. Continuity of satellite measurements (i.e., elimination of gaps in the long-term record) through appropriate launch and orbital strategies should be ensured
14. Rigorous prelaunch instrument characterization and calibration, including radiance confirmation against an international radiance scale provided by a national metrology institute, should be ensured
15. Onboard calibration adequate for climate system observations should be ensured and associated instrument characteristics monitored
16. Operational production of priority climate products should be sustained and peer-reviewed new products should be introduced as appropriate
17. Data systems needed to facilitate user access to climate products, metadata and raw data, including key data for delayed-mode analysis, should be established and maintained
18. Use of functioning baseline instruments that meet the calibration and stability requirements stated above should be maintained for as long as possible, even when these exist on decommissioned satellites
19. Complementary in situ baseline observations for satellite measurements should be maintained through appropriate activities and cooperation
20. Random errors and time-dependent biases in satellite observations and derived products should be identified

was also made available for open review by the community (GCOS, 2009).

The progress assessment for all 131 actions identified in the GCOS-IP showed that good to moderate progress had been achieved for the majority of actions. Nevertheless, in 11 % of the actions, little or no progress could be reported. The report showed evidence that developed countries had improved many of their climate observation capabilities, although commitment to sustained long-term operation remained to be secured for several important observing systems. Developing countries, however, had made only limited progress in filling gaps in their in situ observing networks, with some evidence of decline in some regions. Satellite agencies had improved both mission continuity and observational capability and were increasingly meeting the needs for data reprocessing, product generation, and access. Overall, the global climate observing system had progressed significantly, but still fell short of meeting all the climate information needs of

the UNFCCC and broader user communities at regional and national levels.

As a result, a 2010 update of the “Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC” (GCOS, 2010) was prepared in response to a request by Parties to the UNFCCC expressed at the 30th session of the UNFCCC Subsidiary Body for Scientific and Technological Advice (SBSTA) in June 2009 and confirmed in UNFCCC Decision 9/CP.15 (December 2009). This updated GCOS-IP includes 138 actions and takes account of the latest status of observing systems, recent progress in science and technology, the increased focus on adaptation, enhanced efforts to optimize mitigation measures, and the need for improved predictions of climate change.

The supplemental details to the satellite-based component of the 2010 updated GCOS-IP were publicly reviewed, and comments were taken into account after consultation with the broader GCOS expert community.

The so-called Satellite Supplement (GCOS, 2011) provides additional technical detail to the actions and needs identified in the 2010 updated GCOS-IP related to satellite-based observations for climate for each of the essential climate variables. In particular, it details the specific satellite data records that should be sustained in accordance with the GCOS Climate Monitoring Principles, as well as other important supplemental satellite observations that are needed on occasion or at regular intervals.

Summary

Established in 1992, GCOS is a joint undertaking of the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational Scientific and Cultural Organization (UNESCO), the United Nations Environment Programme (UNEP), and the International Council for Science (ICSU). Its goal is to provide comprehensive information on the total climate system, involving a multidisciplinary range of physical, chemical, and biological properties and atmospheric, oceanic, hydrological, cryospheric, and terrestrial processes. GCOS is intended to be a long-term, user-driven operational system capable of providing the comprehensive observations required for monitoring the climate system, detecting and attributing climate change, assessing impacts of, and supporting adaptation to, climate variability and change, as well as for application to national economic development and for research to improve understanding, modeling, and prediction of the climate system.

The increasing profile of climate change has reinforced worldwide awareness of the importance of an effective global climate observing system. This system has significantly improved over the past years, but much remains to be done to meet the needs of science and society. The consequence of not meeting these requirements would be to seriously compromise the information on, and predictions of, climate variability and change.

Bibliography

- GCOS, 2003. *Second Report on the Adequacy of the Global Observing System for Climate in Support of the UNFCCC (GCOS-82)*, WMO/TD 1143. Geneva: World Meteorological Organisation.
- GCOS, 2004. *Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (GCOS-92)*, WMO/TD 1219. Geneva: World Meteorological Organisation.
- GCOS, 2006. *Systematic Observation Requirements for Satellite-Based Products for Climate – Supplemental Details to the Satellite-Based Component of the Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (GCOS-107)*. Geneva: World Meteorological Organisation.
- GCOS, 2009. *Progress Report on the Implementation of the Global Observing System for Climate in Support of the UNFCCC 2004–2008 (GCOS-129)*. Geneva: World Meteorological Organisation.
- GCOS, 2010. *Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 update) (GCOS-138)*, WMO/TD 1523. Geneva: World Meteorological Organisation.

GCOS, 2011. *Systematic Observation Requirements for Satellite-Based Data Products for Climate (2011 Update), Supplemental Details to the Satellite-Based Component of the “Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update)” (GCOS-154)*. Geneva: World Meteorological Organisation.

Cross-references

Cryosphere, Measurements and Applications
 Global Earth Observation System of Systems (GEOSS)
 Global Land Observing System
 Ocean, Measurements and Applications

GLOBAL EARTH OBSERVATION SYSTEM OF SYSTEMS (GEOSS)

Steffen Fritz
 International Institute for Applied Systems Analysis,
 Laxenburg, Austria

Definition

Global Earth Observation System of Systems (GEOSS) aims to involve all countries of the world to integrate ground-based (in situ), airborne, and space-based observation networks. Those Earth observation (EO) systems which participate in GEOSS retain their existing mandates but share primary observational data as well as information derived from those observations. The sharing of and access to data is enabled through common data standards. GEOSS is designed to address nine societal benefit areas, namely, ecosystems, biodiversity, health, disasters, energy, climate, weather, water, and agriculture. GEOSS seeks to connect the producers of environmental data and decision-support tools with the end users of these products, with the aim of enhancing the relevance of Earth observations to global issues. The end result is a global public infrastructure that generates comprehensive, near-real-time environmental data, information, and analyses for a wide range of users. The Group on Earth Observations (GEO www.earthobservations.org) is coordinating efforts to build GEOSS on the basis of a 10 Year Implementation Plan (GEO, 2005) running from 2005 to 2015. GEO is a voluntary partnership of governments and international organizations. GEO meets in plenary at the senior official level and periodically at the ministerial level. Decisions are taken via consensus of GEO members.

Introduction

We are currently faced with major challenges due to the powerful processes which drive global change. These processes operate at a global scale and can only be observed, understood, and predicted by a system that operates at a supranational level. These processes can have important consequences for human well-being, and the monitoring of these processes is critical in order to understand the

complex system of the Earth's terrestrial and maritime biospheres.

It is envisaged that through GEOSS, the information available to decision makers at all levels will be improved, specifically relating to human health and safety, protection of the global environment, the reduction of losses from natural disasters, and achieving sustainable development. GEOSS is founded on the principle that better international cooperation in the collection, interpretation, and sharing of EO information is an important and cost-effective mechanism for achieving this aim (Fritz et al., 2008). GEOSS will yield a broad range of societal benefits, notably:

1. Reducing loss of life and property from natural and human-induced disasters
2. Understanding environmental factors affecting human health and well-being
3. Improving the management of energy resources
4. Understanding, predicting, mitigating, and adapting to climate variability and change
5. Improving water resource management via better understanding of the water cycle
6. Improving weather information, forecasting, and warning
7. Improving management and protection of terrestrial, coastal, and marine ecosystems
8. Supporting sustainable agriculture and combating desertification
9. Understanding, monitoring, and conserving biodiversity

These correspond to the nine societal benefit areas (SBAs) referred to as disaster, health, energy, climate, water, weather, ecosystems, agriculture, and biodiversity.

Origin of GEOSS

As a result of the Earth Summit in Brazil in 1992, Agenda 21 identified the bridging of the gap between data collection and information required by decision makers as a key priority. As a result of the Summit, three different observing systems were formed, namely, the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS), and the Global Terrestrial Observing System (GTOS). The 2002 World Summit on Sustainable Development in Johannesburg highlighted the urgent need for coordinated observations relating to the state of the Earth. It was realized that only by linking and coordinating the current observing systems could complex Earth processes (in an increasingly environmentally stressed world) be understood (GEO, 2005). The First Earth Observation Summit convened in Washington, D.C., in July 2003, adopted a Declaration establishing the ad hoc intergovernmental Group on Earth Observations (ad hoc GEO) to draft a 10 Year Implementation Plan. The Second Earth Observation Summit in Tokyo, Japan, in April 2004 adopted a Framework Document defining the scope and intent of a Global Earth Observation System of

Systems (GEOSS). The Third Earth Observation Summit, held in Brussels in February 2005, endorsed the GEOSS 10 Year Implementation Plan and established the intergovernmental Group on Earth Observations (GEO) to carry it out (Plag, 2006).

The Group on Earth Observation (GEO)

GEO is building GEOSS on the basis of a 10 Year Implementation Plan (GEO, 2005). The GEO member governments and participating organizations are supported by the GEO secretariat based in Geneva, Switzerland. The secretariat consists of a director appointed by the executive committee, several international civil servants, and 8–10 national technical and scientific experts who are seconded to the secretariat for 2 or 3 years. The secretariat is responsible for coordinating the tasks and other activities that are driving the 10 Year Implementation Plan for GEOSS. The secretariat also services the plenary and the committees and implements outreach and other support activities.

Earth observation systems

EO systems consist of instruments and models designed to measure, monitor, and predict the physical, chemical, and biological aspects of the Earth system. Buoys floating in the oceans monitor temperature and salinity; meteorological stations and balloons record air quality and rainwater trends; sonar and radar systems estimate fish and bird populations; seismic and Global Positioning System stations record movements in the Earth's crust and interior; some 60-plus high-tech environmental satellites scan the planet from space; powerful computerized models generate simulations and forecasts; and early warning systems issue alerts to vulnerable populations.

These various systems have typically operated in isolation from one another. In recent years, however, sophisticated new technologies for gathering vast quantities of near-real-time and high-resolution EO data have become operational. At the same time, improved forecasting models and decision-support tools are increasingly allowing decision makers and other users of EO to fully exploit this widening stream of information.

With investments in EO now reaching a critical mass, it has become possible to link diverse observing systems together to paint a full picture of the Earth's condition. Because the costs and logistics of expanding EO are daunting for any single nation, linking systems together through international cooperation also offers cost savings.

Implementing GEOSS

As a networked system, GEOSS is owned by all of the GEO members and participating organizations. Partners maintain full control of the components and activities that they contribute to the system of systems. Implementation is being pursued through a work plan (GEO, 2008) currently consisting of over 70 tasks. Each task supports

one of the nine SBAs or four transverse areas and is carried out by interested organizations.

GEO work plans are triennially revised and provide the agreed framework for implementing the GEOSS 10 Year Implementation Plan 2005–2015 (www.earthobservations.org/documents). These work plans consist of a set of practical tasks that are carried out by various GEO members and participating organizations. Connections will be realized between diverse observing, processing, data-assimilation, modeling, and information-dissemination systems. This will make it possible to obtain a considerably increased range of data sets, products, and services on the key aspects of the Earth system. The plans are also focusing on enhancing the role of users and reflect the inputs and engagement of the communities of practice, taking full account of the Integrated Global Observing Strategy (IGOS, www.igospartners.org/) transition into GEO.

Governments and participating organizations have advanced GEOSS by contributing a variety of “Early Achievements.” These “First 100 Steps to GEOSS” (http://www.earthobservations.org/documents/the_first_100_steps_to_geoss.pdf) were presented to the 2007 Cape Town Ministerial Summit.

Worldwide, several parallel initiatives are contributing their data to GEOSS. Inter alia, Europe is establishing Global Monitoring for Environment and Security (GMES www.gmes.info), the USA is building the Integrated Earth Observation System (IOES <http://usgeo.gov/>), and China and Brazil are collaborating through the China-Brazil Earth Resources Satellite Programme (CBERS, www.cbears.inpe.br), which is launching a new Earth observation service that will provide state-of-the-art images of the planet to end users throughout Africa free of charge.

The ultimate objective of GEOSS is to develop the use of EO by a broad range of user communities – from both developed and developing countries and ranging from decision- and policy makers to scientists, industry, international governmental, and nongovernmental organizations. Engagement of these communities to identify their needs for new or improved data is essential to enhance the adequacy of provided services and products for a wide diversity of applications.

Data standards and data dissemination

Due to the fact that EO data is obtained from a multitude of sources, an enormous effort is required among different governments and user groups to achieve true data interoperability (Durbha et al., 2008). Therefore, common standards for architecture and data sharing are essential (see *GEOSS Best Practices WIKI* <http://wiki.ieee-earth.org/> and the GEOSS Standards and Interoperability Registry www.earthobservations.org/gci_sr.shtml). Each contributor to GEOSS must subscribe to the GEO data-sharing principles, which aim to ensure the full and open exchange of data, metadata, and products. These issues are fundamental to the successful operation of GEOSS.

The architecture of an Earth observation system refers to the way in which its components are designed so that they function as a whole. Each GEOSS component must be included in the GEOSS registry and configured so that it can communicate with the other participating systems. The GEOSS Components and Services Registry provides a formal listing and description of all the Earth observation systems, data sets, models, and other services and tools that together constitute the Global Earth Observation System of Systems (www.earthobservations.org/gci_cr.shtml).

GEOSS will disseminate information and analyses directly to users. GEO is developing the GEOPortal (www.geoportal.org) as a single Internet gateway to the data produced by GEOSS. The purpose of GEOPortal is to make it easier to integrate diverse data sets, identify relevant data and portals of contributing systems, and access models and other decision-support tools. For users without good access to high-speed Internet, GEO has established GEONETCast (<http://www.earthobservations.org/geonetcast.shtml>), a system of four communications satellites that transmit data to low-cost receiving stations maintained by the users.

Activities under the SBAs

1. SBA Disasters GEOSS implementation will bring a more timely dissemination of information through better coordinated systems for monitoring, predicting, risk assessment, early warning, mitigating, and responding to hazards at local, national, regional, and global levels (GEO, 2005). Cozannet et al. (2008) describe a prototype catalogue that was developed to improve access to information about sensor networks surveying geological hazards (geohazards). Related Project: for example, GFMC (www.fire.uni-freiburg.de/)
2. SBA Human Health GEOSS will improve the flow of appropriate environmental data and health statistics to the health community, promoting a focus on prevention and contributing to continued improvements in human health worldwide (GEO, 2005). Related Project: for example, PROMOTE (<http://www.gse-promote.org/>)
3. SBA Energy Resources GEOSS outcomes in the energy area will support environmentally responsible and equitable energy management, better matching of energy supply and demand, reduction of risks to energy infrastructure, more accurate inventories of greenhouse gases and pollutants, and a better understanding of renewable energy potential (GEO, 2005). Related Project: for example, ENVISOLAR (<http://www.envisolar.com/>)
4. SBA Climate GEOSS outcomes will enhance the capacity to model, mitigate, and adapt to climate change and variability. Better understanding of the climate and its impacts on the Earth system, including its human and economic aspects, will contribute to improved climate prediction and facilitate sustainable development while avoiding dangerous perturbations

to the climate system (GEO, 2005). A global climate observation system is essential to improve our understanding of the climate system and our ability to anticipate trends (Fellous, 2008). Related Project: for example, APEC (<http://www.apcc21.net/>)

5. SBA Water GEOSS implementation will improve integrated water resource management by bringing together observations, prediction, and decision-support systems and by creating better linkages to climate and other data. In situ networks and the automation of data collection will be consolidated, and the capacity to collect and use hydrological observations will be built where it is lacking (GEO, 2005). Related Project: for example, GRDC (<http://grdc.bafg.de/servlet/is/Entry.987.Display/>)
6. SBA Weather GEOSS can help fill critical gaps in the observation of, for example, wind and humidity profiles, precipitation, and data collection over ocean areas; extend the use of dynamic sampling methods globally; improve the initialization of forecasts; and increase the capacity in developing countries to deliver essential observations and use forecast products. Access to weather data for the other SBAs will be facilitated (GEO, 2005). Related Project: for example, TIGGE (<http://tigge.ecmwf.int/>)
7. SBA Ecosystems GEOSS implementation will seek to ensure that methodologies and observations are available on a global basis to detect and predict changes in ecosystem condition and to define resource potentials and limits. Ecosystem observations will be better harmonized and shared, spatial and topical gaps will be filled, and in situ data will be better integrated with space-based observations. Continuity of observations for monitoring wild fisheries, the carbon and nitrogen cycles, canopy properties, ocean color, and temperature will be set in place (GEO, 2005). This SBA is strongly linked to supporting the monitoring of the state of forests and to provide essential information to the UNFCCC process for REDD activities as well as the monitoring of illegal logging. Related Project: for example, POSTEL (<http://postel.mediasfrance.org>)
8. SBA Agriculture GEOSS implementation will address the continuity of critical data, such as high-resolution observation data from satellites. A truly global mapping and information service, integrating spatially explicit socioeconomic data with agricultural, forest, and aquaculture data will be feasible, with applications in poverty and food monitoring, international planning, and sustainable development (GEO, 2005). Related Project: for example, GAMS (http://www.earthobservations.org/cop_ag_gams.shtml)
9. SBA Biodiversity Implementing GEOSS will unify many disparate biodiversity-observing systems and create a platform to integrate biodiversity data with other types of information. Taxonomic and spatial gaps will be filled, and the pace of information collection and dissemination will be increased (GEO, 2005). Since biodiversity data in general is not lacking, but

often uneven in its spatial, temporal, and topological coverage as well as physically dispersed and unorganized, the GEOBON project tries to overcome these shortcomings by installing a system which aims to organize the information, increase the exchange between suppliers and users, and to create a mechanism whereby data of different kinds, from many sources, can be combined (Scholes et al., 2008). Related Project: for example, GEOBON www.earthobservations.org/cop_bi_geobon.shtml

Conclusions

Without a global effort to link all current observing systems to build the Global Earth Observation System of Systems (GEOSS), modern civilization will be struggling to understand the complex chemical, biological, and physical processes of the Earth system. Therefore, GEOSS is needed more than ever to acquire comprehensive, near-real-time environmental data, information, and analysis by users as well as decision makers to respond more effectively to the plethora of environmental challenges.

Since the establishment of the Group on Earth Observations (GEO), many early achievements have been realized. These are documented in www.earthobservations.org under each of the nine GEOSS themes or societal benefit areas (SBAs). However, the active mobilization of data users and providers will remain necessary to make GEOSS a success (Fellous, 2008).

Many EO resources have been created and are available to the global community in order to support scientists, decision makers, and the general populace. To realize a successful GEOSS, the key is to provide mechanisms that enable EO data and geospatial data from those resources to be processed, shared, and coordinated. To this end, the GEOPortal (www.geoportal.org) has been developed as the Internet gateway to the data produced by GEOSS.

Global Earth Observations (EO) may be instrumental to achieve sustainable development, but to date there have been no integrated assessments of their economic, social, and environmental benefits. The project Global Earth Observation – Benefit Estimation (GeoBene www.geo-bene.eu) is developing methodologies and analytical tools to assess the societal benefits of GEOSS. First results from the GeoBene project illustrate that the overall societal benefit is by far higher than the incremental costs necessary to establish GEOSS.

Bibliography

- Cozannet, G. L., Hosford, S., Douglas, J., Serrano, J. J., Coraboeuf, D., and Comte, J., 2008. Connecting hazard analysts and risk managers to sensor information. *Sensors*, **8**, 3932–3937.
- Durbha, S. S., King, R. L., and Younan, N. H., 2008. An information semantics approach for knowledge management and interoperability for the global earth observation systems of systems. *IEEE Systems Journal*, **2**, 358–365.
- Fellous, J. L., 2008. Towards a global climate observing system. *Interdisciplinary Science Reviews*, **3**, 83–94.

- Fritz, S., Scholes, R. J., Obersteiner, M., Bouma, J., and Reyers, B., 2008. A conceptual framework for assessing the benefits of a global earth observation system of systems. *IEEE Systems Journal*, 2, 338–348.
- GEO, 2005. *Global Earth Observation System of Systems: 10-Year Implementation Plan Reference Document*. ESA, www.earthobservations.org/documents, 209 pp.
- GEO, 2008. *GEO 2007–2009 Work Plan Toward Convergence*. GEO, www.earthobservations.org/documents/wp0709_v6.pdf, 30 pp.
- http://nng.esoc.esa.de/ws2006/Papers/01_Plug_geo_pospaper_11pt_new.pdf.gz.
- Plag, H. P., 2006. *Geo, Geoss and Igos-P: The Framework of Global Earth Observations*. ftp://igscb.jpl.nasa.gov/pub/resource/pubs/06_darmstadt/IGS%20WS%202006%20Papers%20PDF/1_Plug_geo_pospaper_11pt_new.pdf
- Scholes, R. J., Mace, J. M., Turner, W., Geller, G. N., Jürgens, N., Larigauderie, A., Muchoney, D., Walther, B. A., and Mooney, H. A., 2008. Towards a global biodiversity observing system. *Science*, 321, 1044–1045.

Cross-references

Cryosphere and Polar Region Observing System
Global Climate Observing System
Global Land Observing System

GLOBAL LAND OBSERVING SYSTEM

Johannes A. Dolman
Department of Earth Sciences, VU University
Amsterdam, Amsterdam, The Netherlands

Definition

An integrated system of in situ and remote sensing observation that provides information about the state of the global land surface.

Introduction

Land has a wide variety of natural features, slopes, vegetation, and soils that affect water budgets, carbon fluxes, and the reflective properties of the surface. Land is often covered by vegetation; importantly, almost 40 % of the Earth's land surface is now under some form of management. Land use changes the characteristics of the land surface and thus can induce important local climate effects, especially through changes in albedo, roughness, soil moisture, and evapotranspiration. Precise quantification of the rate of change is important to determine whether feedback or amplification mechanisms are operating through terrestrial processes to affect the climate system.

Humans have long history of observing state of the land surface through in situ observations. This relates not only to the earliest instrumental observations of surface air temperature by Buys Ballot, but also travelers generated extended descriptions of the state of the land surface they encountered on their travels. The advance of satellite technology has made important breakthroughs in our

observing capability, as now for the first time in history we are able to observe at global level the state of the Earth's surface. In 1983, a group of around 100 scientists met to initiate the International Satellite Land Surface Climatology Project (ISLSCP). This project was set out (Seller and Hall, 1992) to (1) monitor global-scale fluctuations of the land surface caused by human interference or climatic fluctuations, (2) further develop mathematical models designed to predict or simulate climate on various time scales, and (3) permit inclusion of land surface climatological variables in diagnostic and empirical studies of climatological variations. In 1984, CEOS (Committee on Earth Observation Satellites) was established in response to a recommendation from a Panel of Experts on Remote Sensing from Space, under the aegis of the G7 Economic Summit. This group, recognizing the multidisciplinary nature of satellite Earth observation, aims to coordinate satellite missions between the various space agencies, particularly with a view to avoid the occurrence of gaps. For the land surface, there is a specific constellation (<http://wgiss.ceos.org/lcip/>) for Land Surface Imaging established. There are three ways to observe the state of the land surface through remote sensing, in the optical, thermal, and microwave frequencies.

Microwave sensors

Active microwave instruments (radars) transmit at frequencies of around 1–10 GHz and measure the backscattered signals to generate microwave images of Earth's surface at high spatial resolutions (between 1 and 100 m) and with a swath width of around 100 km. The images produced have a similar resolution to those from high-resolution optical imagers, but radars have the capability to “see” through clouds providing data on all weather, on day/night basis. This is particularly useful to assess tropical deforestation. SAR interferometry can record the phase shift between two SAR images recorded at slightly different times, thereby providing highly accurate information on the motion of surfaces. Examples of such systems are the ASAR on ESA's ENVISAT platform, PALSAR (JAXA), and RADARSAT. An important application is also the estimation of surface soil moisture at high resolution (1 km) from active radar systems (Wagner et al., 2008) and the behavior of ice (Rignot and Kanagaratnam, 2006). Passive microwave techniques receive the microwave signal emitted from the Earth's surface. This signal can be used to retrieve soil moisture (Owe et al., 2008), surface temperature (Holmes et al., 2009), vegetation water content (Shi et al., 2008), and snow characteristics (Chang et al., 1987) at relatively large scales. The Soil Moisture and Salinity Mission from ESA (SMOS) has been successfully launched in 2009, while NASA's SMAP (Soil Moisture Active and Passive) is planned for launch in 2015. The use of P-band Synthetic Aperture Radar (SAR) is currently investigated to assess its capability to observe forest biomass from space.

Optical and thermal infrared sensors

Mid-resolution optical satellite systems image the Earth's land surface in visible, near-infrared, shortwave infrared, and thermal infrared wavelengths with spatial resolutions between 10 and 100 m. Examples include the Landsat series, the first launched in 1972, with now Landsat 7 nearing its final operational time. In 2008, the US Geological Survey made the Landsat archives – which extend back to 1972 – freely available. These data are being used to measure rates and patterns of land cover change, first on a continental and pan-tropical scale, and eventually globally. Other satellites include the Japanese ADEOS satellite, the French SPOT, and the Brazilian-Chinese CBERS satellite. High-resolution imagery is provided at 1 m resolution by the IKONOS satellite, while QUICKBIRD can achieve resolutions better than 1 m. Optical sensors are increasingly used to study fundamental properties of the land surface such as fraction of photosynthetically active radiation and Leaf Area Index (LAI). These properties can be used to assess physiological changes in the land surface and as such go beyond the more classical land use classifications (e.g., Gobron and Belward, 2009). The thermal infrared sensors have mainly been used to assess hydrometeorological state variables and fluxes (e.g., Norman et al., 1995) and geological mapping (Kahle and Goetz, 1983).

Future and new developments

Europe's first ice mission, CryoSat-2 carries a precise radar altimeter to measure changes at the margins of the ice sheets that overlay Greenland and Antarctica. By measuring thickness change in ice, CryoSat-2 will provide information on the stability of the Earth's ice sheets.

Among promising new technologies in land surface remote sensing is lidar (Dubayah et al., 2010). Spaceborne lidar has the potential to retrieve many aspects of forest structure important for carbon and ecosystem studies, including canopy height, leaf distribution, and aboveground biomass stocks. The ICESAT (Ice, Cloud, and land Elevation Satellite) mission is currently in orbit, while the planned DESDynI (Deformation, Ecosystem Structure, and Dynamics of Ice) mission will combine a multi-beam lidar with polarimetric and interferometric SAR capability to measure forest structure, biomass, and their dynamics.

Integrated global land observations

Foundations for both in situ observation networks and space-based observing components for the land surface are in place, and continuity of missions and validation still need to be strengthened. Improvements in understanding of the terrestrial components of the climate system, the causes and response of this system to change, and consequences in terms of impact and adaptation are vital to society. Increasing significance is thus being placed on terrestrial data for estimating climate forcing and better understanding of climate change and variability, as well as for impact and mitigation assessment. Recognition of

this need makes integration of remote sensing-based and in situ data a priority issue in global change research.

Summary

Land has a wide variety of natural features, slopes, vegetation, and soils that affect water budgets, carbon fluxes, and the reflective properties of the surface. To be able to observe the properties of the land surface, both in situ and remote sensing tools are needed. There are three ways to observe the state of the land surface through remote sensing, in the optical, thermal, and microwave frequencies. We briefly review the most important land applications in these domains.

We conclude that foundations exist for both the in situ observation networks and space-based observing components, but that continuity of missions and validation still need to be strengthened. Improvements in understanding of the terrestrial components of the climate system, the causes and response of this system to change, and consequences in terms of impact and adaptation are vital to society.

Bibliography

- Chang, A., Foster, J., and Hall, D., 1987. Nimbus-7 SMMR derived globals now cover parameters. *Annals of Glaciology*, **9**, 39–44.
- Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., and Chazdon, R. L., 2010. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research*, **115**, G00E09, doi:10.1029/2009JG000933.
- Gobron, N., and Belward, A., 2009. Global vegetation condition, in state of the climate in 2008. *Bulletin of the American Meteorological Society*, **90**, S44–S45.
- Holmes, T. R. H., De Jeu, R. A. M., Owe, M., and Dolman, A. J., 2009. Land surface temperatures from Ka-Band (37 GHz) passive microwave observations. *Journal of Geophysical Research*, **114**, D04113, doi:10.1029/2008JD010257.
- Kahle, A., and Goetz, A., 1983. Mineralogic information from a new airborne thermal infrared multispectral scanner. *Science*, **222**, 24–27.
- Norman, J., Kustas, W., and Humes, K., 1995. A two source approach for estimating soil and vegetation energy fluxes from observations of directional radiometric surface temperature. *Agricultural and Forest Meteorology*, **77**, 263–293.
- Owe, M., De Jeu, R. A. M., and Holmes, T. R. H., 2008. Multi-sensor historical climatology of satellite-derived global land surface moisture. *Journal of Geophysical Research*, **113**, F01002, doi:10.1029/2007JF000769.
- Rignot, E., and Kanagaratnam, P., 2006. Changes in the velocity structure of the Greenland ice sheet. *Science*, **311**(5763), 986–990.
- Seller, P. J., and Hall, F. G., 1992. FIFE in 1992: Results, scientific gains, and future research directions. *J. Geophys. Res.*, **97**, 19091–19109.
- Shi, J., Jackson, T., Tao, J., Du, J., Blindish, R., Lu, L., and Chen, K. S., 2008. Microwave vegetation indices for short vegetation covers from satellite passive microwave sensor AMSR-E. *Remote Sensing of Environment*, **112**(2008), 4285–4300.
- Wagner, W., Pathe, C., Doubkova, M., Sabel, D., Bartsch, A., Hasenauer, S., Blöschl, G., Scipal, K., Martínez-Fernández, J., and Löw, A., 2008. Temporal stability of soil moisture and radar backscatter observed by the advanced synthetic aperture radar (ASAR). *Sensors*, **8**, 1174–1197.

GLOBAL PROGRAMS, OPERATIONAL SYSTEMS

Mary Kicza

National Oceanic and Atmospheric Administration
(NOAA), Washington, DC, USA

Definition

Global programs, operational systems are international coordination mechanisms that have been established to coordinate the world's operational Earth remote sensing activities.

Introduction

Understanding the global environment requires an international effort: The mission objectives of Earth observation systems overlap, remote sensing science is conducted in universities and research centers around the world, and remotely sensed data have interdisciplinary applications. A number of international coordination mechanisms have been established to manage the world's operational Earth remote sensing activities. They address issues ranging from satellite system design and deployment to data acquisition and dissemination. Some of the major mechanisms are summarized below.

Asia-Pacific satellite data exchange and utilization group (APSDEU)

The Asia-Pacific Satellite Data Exchange and Utilization Group was established to increase the amount of data exchanged among agencies in the Asia-Pacific region over existing communications links, to apprise regional agencies of changes and improvements to communications and satellite systems, and to identify means of improving existing data utilization. Participants include the Japanese Meteorological Agency, the China Meteorological Administration, the Korean Meteorological Administration, the Australia Bureau of Meteorology, the Hong Kong Observatory, the Canadian Meteorological Centre, and the United States National Oceanic and Atmospheric Administration.

Committee on Earth observation satellites (CEOS)

CEOS was established in 1984 at the request of the Economic Summit of Industrialised Nations Working Group (G7) as the international forum for Earth observing space agencies. CEOS contributes the space component of the Group on Earth Observation's Global Earth Observation System of Systems and supports key stakeholders with a wide range of Earth observation data, products, and expertise.

CEOS coordinates civil spaceborne observations of the Earth and operates through the best efforts of its 30 members (space agencies) and 22 associates via voluntary contributions. Participating agencies strive to address critical scientific questions and to plan satellite missions without unnecessary overlap. CEOS has three primary objectives in pursuing this goal: (1) to optimize benefits of

spaceborne Earth observations through cooperation of its members in mission planning and in development of compatible data products, formats, services, applications, and policies; (2) to serve as a focal point for international coordination of space-related Earth observation activities; and (3) to exchange policy and technical information to encourage complementarity and compatibility of observation and data exchange systems.

CEOS is composed of four main implementation mechanisms: the CEOS Strategic Implementation Team (SIT); a permanent CEOS Secretariat; four Working Groups; and seven CEOS Virtual Constellations for GEO.

Coordination group for meteorological satellites (CGMS)

CGMS was formed in 1972 and currently has 19 member agencies and 7 observer agencies. CGMS provides a forum for the exchange of technical information on geostationary and polar-orbiting meteorological satellite systems, as well as on research and development satellite missions. CGMS works to harmonize meteorological satellite mission parameters, such as orbits, sensors, data formats, and downlink frequencies. CGMS also encourages compatibility and mutual backup for system failures through cooperative mission planning, compatible data products and services, and the coordination of space and data-related activities, thereby complementing the work of other international satellite coordinating mechanisms.

Global climate observing system (GCOS)

The Global Climate Observing System was established in 1992 as a result of the second World Climate Conference in 1990. GCOS is cosponsored by the World Meteorological Organization, the Intergovernmental Oceanographic Commission, the United Nations Environment Programme, and the International Council for Science and provides a framework for meeting worldwide requirements for a sustained and robust global system of climate observations from atmospheric, oceanic, terrestrial, and remote sensing platforms. GCOS constitutes the climate observing component of the Group on Earth Observation's Global Earth Observation System of Systems.

While GCOS neither makes observations nor generates data products itself, it stimulates, encourages, coordinates, and otherwise facilitates the taking of the needed observations by national or international organizations. GCOS provides an operational framework for integrating the observational systems of participating countries and organizations into a comprehensive climate observing system. As such, GCOS serves as the formal observations compliance mechanism for the United Nations Framework Convention on Climate Change and depends on a network of national GCOS focal points, scientists, and a multitude of operational and research agencies to implement its requirements for observing the GCOS Essential Climate Variables (ECV).

Group on Earth observations (GEO)

GEO is coordinating international efforts to build a Global Earth Observation System of Systems (GEOSS). This emerging public infrastructure is interconnecting a diverse and growing array of instruments and systems for monitoring and forecasting changes in the global environment. This "system of systems" supports policymakers, resource managers, science researchers, and many other experts and decision makers.

GEO was launched in response to calls for action by the 2002 World Summit on Sustainable Development and by the G8 (Group of Eight) leading industrialized countries. These high-level meetings recognized that international collaboration is essential for exploiting the growing potential of Earth observations to support decision making in an increasingly complex and environmentally stressed world.

GEO is a voluntary partnership of governments and international organizations. It provides a framework within which these partners can develop new projects and coordinate their strategies and investments. As of March 2012, GEO's Members included 88 governments and the European Commission. In addition, 64 intergovernmental, international, and regional organizations with a mandate in Earth observation or related issues have been recognized as Participating Organizations.

International charter "space and major disasters"

The International Charter "Space and Major Disasters" was initiated at the UNISPACE III conference in Vienna in 1999. The founding members were the European Space Agency, the French Centre National d'Etudes Spatiales, and the Canadian Space Agency.

The Charter represents a joint effort by global space agencies to put Earth observation resources at the service of rescue authorities responding to major disasters. Recognizing that no single satellite operator can address all the challenges associated with disaster management, each member agency commits resources to support the provisions of the Charter and to help mitigate the effects of disasters on human life and property.

North America-Europe data exchange group (NAEDX)

The North America-Europe Data Exchange Group was established in 1988 to optimize, maintain, and improve the exchange of meteorological data and products between Europe and North America; to maintain and update consolidated European and North American user requirements; and to provide a forum for discussions on planned changes and problems in data dissemination, processing, and formats, as well as optimization of telecommunications and IT infrastructure. Participants include Deutscher Wetterdienst, the European Center for Medium-Range Weather Forecasts, the United Kingdom Met Office, the Meteorological Service of Canada, Météo-France, the United States National Oceanic and Atmospheric Administration, and the European Organisation for the Exploitation of Meteorological Satellites.

World meteorological organization space programme (WMOSP)

The World Meteorological Organization Space Programme coordinates environmental satellite matters and activities throughout all WMO Programmes and provides guidance on the potential of remote sensing techniques in meteorology, hydrology and related disciplines, and applications. Its work includes the optimization of international space-based Earth observing capabilities with the goal of enhancing the space-based component of WMO's World Weather Watch Global Observing System.

Summary

Understanding the global environment requires an international effort. A number of international coordination mechanisms have been established to coordinate the world's operational Earth remote sensing activities that address issues ranging from satellite system design and deployment to data acquisition and dissemination. These include the Asia-Pacific Satellite Data Exchange and Utilization Group, the Committee on Earth Observation Satellites, the Coordination Group for Meteorological Satellites, the Global Climate Observing System, the Group on Earth Observations, the International Charter "Space and Major Disasters," the North America-Europe Data Exchange Group, and the World Meteorological Organization Space Programme.

Bibliography

<http://gcoss.wmo.int/>
<http://gosis.org>
<http://www.ceos.org/>
<http://www.cgms-info.org/>
<http://www.disasterscharter.org/>
<http://www.earthobservations.org/>

Cross-references

Data Access
 Data Policies
 Global Climate Observing System
 Global Earth Observation System of Systems (GEOSS)
 Global Land Observing System
 International Collaboration

GPS, OCCULTATION SYSTEMS

Chi O. Ao
 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

GNSS (Global Navigation Satellite System) occultation; GNSS Radio Occultation (GNSS RO); GPS Radio Occultation (GPS RO or GPSRO)

Definition

GPS/GNSS Radio Occultation. An atmospheric profiling technique whereby the amplitude and phase delay of the radio signals from the global positioning satellites are observed across the Earth's limb as the satellites rise or set with respect to the receiving platform.

Introduction

A Global Positioning System (GPS) radio occultation (RO) system consists of a GPS receiver on a low-Earth-orbit spacecraft that is capable of tracking the radio signals transmitted by one of the GPS satellites as it rises or sets through the Earth's atmosphere. For each occultation event, the recorded measurements consist of the time series of the carrier phase and the amplitude for each of the two GPS frequencies. With precise knowledge of orbits and clocks for the transmitter and receiver, these measurements can be directly inverted to derive a vertical profile of refractive index n (or refractivity $N = (n - 1) \times 10^6$) from the surface of the Earth to the ionosphere. In the neutral atmosphere, the refractive index can be used to retrieve atmospheric pressure, temperature, and tropospheric humidity profiles with additional constraints. In the ionosphere, electron density profiles can be derived from the refractive index (Figure 1).

The GPS occultation technique has several unique properties compared to other remote sensing instruments. First, it is a coherent, active limb sounding technique that is capable of yielding atmospheric profiles with very high vertical resolution (from 100 m in the lower troposphere to about 1 km up to the middle stratosphere). Second, the GPS RO measurements are "self-calibrating," meaning that they do not need to be adjusted against external "standard" references that could drift over time. This makes the GPS occultation a valuable dataset for climate benchmarking. Third, the GPS signal frequencies are in the L-band and, therefore, essentially unaffected by the presence of clouds and precipitation. This ensures that the GPS occultation measurements will not be degraded or biased in cloudy regions or under severe weather conditions.

Instrumentation, missions, and data coverage

The RO technique was first developed to probe planetary atmospheres starting in the 1960s. The availability of freely accessible radio signals from the GPS constellation made this concept very appealing for Earth remote sensing in terms of cost-effectiveness and scientific merits (Gurvich and Krasil'nikova 1987; Yunck et al., 1988).

The first GPS RO mission was the proof-of-concept GPS/Meteorology (GPS/MET) that operated in 1995–1997 (Ware et al., 1996). It was equipped with the NASA/JPL TurboRogue GPS receiver modified to acquire and track occultation signals. The success of GPS/MET led to the inclusion of the more advanced "BlackJack" GPS RO receivers on the CHAMP (CHALLENGING Minisatellite Payload) (Wickert et al., 2001) and SAC-C (Satélite de Aplicaciones Científicas-C)

(Hajj et al., 2004) satellites, which have produced GPS RO data nearly continuously from 2001 to 2008. The twin satellites of Gravity Recovery And Climate Experiment (GRACE) also carried the BlackJack receivers, and routine occultation measurements have been made since 2006. The most dramatic addition to the suite of GPS occultation sensors in space is the six-spacecraft COSMIC (Constellation Observing System for Meteorology, Ionosphere, and Climate, also known as FORMOSAT-3) constellation launched in April 2006 (Anthes et al., 2008), which more than double the existing number of occultation soundings observed each day. The COSMIC spacecraft are equipped with the IGOR (Integrated GPS Occultation Receiver), which is based on the BlackJack design. The BlackJack/IGOR can be configured to track the GPS signal in open loop instead of the traditional phase-locked loop. Open-loop tracking is necessary for tracking the rising occultation and is essential for accurately tracking the highly dynamic signals through the moist lower troposphere (Sokolovskiy et al., 2006; Ao et al., 2009).

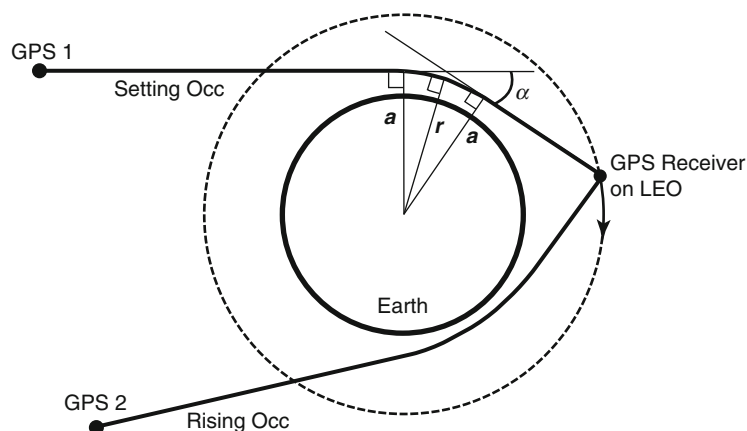
In addition, the ESA/EUMETSAT weather satellites MetOp are equipped with the GRAS GPS occultation receivers. The first of the series, MetOp-A, was launched in October 2006 and have been producing quality occultation data soon afterward (von Engeln et al., 2009). Other new missions currently flying (e.g., TerraSAR-X, TanDEM-X, MetOp-B, OCEANSAT-2) or in planning promise continued availability of GPS RO measurements.

The missions mentioned above are primarily polar orbit satellites in low Earth orbits. Some of the satellites are equipped with both fore- and aft-viewing antennas and could perform occultation measurements in both setting and rising configurations. Approximately 250 occultations, distributed quasi-randomly around the globe, can be obtained per antenna per satellite per day.

Atmospheric retrieval

There are four basic steps in retrieving atmospheric profiles from the GPS RO measurements: (1) calculation of bending angles, (2) removal of ionospheric effects, (3) calculation of refractivity profile, and (4) derivation of the temperature, pressure, and humidity profiles.

The bending angle profile represents a ray optics description of the GPS RO measurements. Traditionally, it is calculated from the Doppler shift of the signal, which is proportional to the rate of change of the phase (Fjeldbo et al., 1971). The "Doppler" method works well under most circumstances; however, it has two shortcomings. First, it is based on geometric optics and hence could not resolve vertical structures smaller than a Fresnel diameter (approximately 1 km in the upper troposphere and lower stratosphere). In addition, the Doppler method assumes that only one ray reaches the receiver at each time. In the moist lower troposphere, atmospheric multipaths occur frequently due to the presence of strong vertical gradients in the refractive index (Sokolovskiy, 2001). To untangle the atmospheric multipaths, diffraction-based



GPS, Occultation Systems, Figure 1 Illustration of GPS RO geometry. A GPS receiver on the low Earth orbit (LEO) measures the changes in the GPS signals as the GPS satellite rises or sets. Due to the vertical variation of the refractivity, the GPS signals are bent and are characterized by the bending angle α and impact parameter a .

methods must be used to transform the received signal from the time domain to impact parameter domain using the Canonical Transform Method (Gorbunov, 2002) or to Doppler frequency domain using the Full Spectrum Inversion (Jensen et al., 2003).

The bending angles are computed for each GPS frequency. Because the GPS signals traverse the ionosphere on its way to the receiver, the bending angles contain contributions from both the ionosphere and the neutral atmosphere. Thus, it is important to remove the ionospheric effects from the bending angles. Due to the dispersive nature of the ionosphere, the first-order (inverse frequency-squared) ionospheric effects can be removed through a linear combination of the dual-frequency bending angles (Vorob'ev and Krasil'nikova, 1994).

The ionosphere-corrected bending angle profile can now be used to compute the refractivity profile directly using the Abel inversion integral (Fjeldbo et al., 1971). Due to the exponentially decreasing atmospheric density with altitude, the bending angles in the upper stratosphere or above are small and susceptible to measurement noise. To reduce the noise in the retrieved refractivity, it is often necessary to replace the bending angles at high altitudes with simple approximations. This can introduce a significant source of uncertainty in stratospheric retrievals (Kursinski et al., 1997; Ao et al., 2006).

Finally, temperature, pressure, and water vapor profiles are derived from the refractivity profile through the use of the microwave refractivity equation (Smith and Weintraub, 1953) and the hydrostatic equation. When water vapor can be neglected, the refractivity is proportional to air density; the temperature and pressure can be determined unambiguously except for the initialization of the hydrostatic equation. At low altitudes, it is not possible to solve simultaneously for these variables without additional constraint. There exist two common

approaches. The first is to assume that temperature is known and solve for pressure and water vapor pressure (e.g., Hajj et al., 2002). The second is to use a one-dimensional variational (1-D Var) technique that provides an optimal estimation of the all these variables assuming known error characteristics of the refractivity measurements and the a priori background fields (Healy and Eyre, 2000).

The accuracy and precision of the GPS RO retrievals have been well studied (Kursinski et al., 1997; Kuo et al., 2004). Generally speaking, the GPS RO refractivity and temperature retrievals are most accurate in the altitude range of 8–25 km (upper troposphere and lower stratosphere, or UTLS), with refractivity error of 0.2 % and temperature error less than 1 K. Above 25 km, the ray bending is small so that the retrieval is susceptible to instrument noise and other systematic errors. Below 8 km, the horizontal structure of the atmosphere becomes a significant error source, leading to a refractivity error of 1–3 %.

Comparisons between GPS RO retrievals with collocated radiosonde sounding, global weather analyses, as well as infrared and microwave satellite measurements have shown that the standard deviations in UTLS are about 1 % in refractivity and 1.5 K in temperature. Analysis of collocated CHAMP and SAC-C occultations found them to be consistent to 0.1 K in the mean and 1 K in standard deviation in UTLS (Hajj et al., 2004). Even better agreement can be found among collocated COSMIC occultations, which are shown to have a standard deviation of 0.2 % in refractivity in UTLS (Schreiner et al., 2007).

Applications

GPS occultation measurements have been used to improve weather forecasts and to improve our understanding of the atmosphere and climate.

Recognizing the unique values of GPS occultation, a number of weather centers around the world have been ingesting GPS bending angle and/or refractivity measurements into their operational numerical weather prediction (NWP) models. Past studies have shown that assimilation of GPS occultation data yielded positive impacts on the forecasts (e.g., Healy and Thépaut, 2006; Cucurull et al., 2008). A key advantage of assimilating GPS occultation is that the measurements can be assimilated into the NWP models without a bias correction and thus can serve as important anchor for assimilating other measurements that do require bias correction (Poli et al. 2010).

One of the most promising applications of GPS occultation is its potential for climate benchmarking (Goody et al., 1998). Detecting small long-term climate change signals requires precise measurements that do not vary as a result of changes in instrumentation over time. The self-calibrating nature of the GPS occultation measurements means that the measurements are not subject to biases between different satellites or time-dependent drifts due to orbit changes, problems that have plagued the efforts to infer long-term temperature trends from the Microwave Sounding Unit (MSU) (Schröder et al., 2003; Steiner et al., 2007).

Another scientific area where GPS occultation has proven useful is in delineating the characteristics of the tropopause, which separates the convectively mixed troposphere (where temperature decreases with height) and the convectively stable stratosphere (where temperature increases with height). The tropopause plays a crucial role in tropical dynamics and the vertical transport of trace gases; moreover, the tropopause height can be a sensitive indicator of climate change. The high vertical resolution that GPS occultation temperature profiles makes them especially suitable for studying the tropopause (Nishida et al., 2000; Randel et al., 2003; Schmidt et al., 2004; Son et al., 2011).

There has also been active research in using GPS occultation data to infer the height of the planetary boundary layer (PBL) (Sokolovskiy et al., 2006; Basha and Ratnam, 2009; Guo et al., 2011; Ao et al., 2012). The PBL height is a key parameter that describes much of the diurnal, synoptic, and climatological processes associated with the PBL in a given region, including its cloud characterization and connections between the surface and free troposphere. It is, however, poorly determined due to a lack of global observations. The fine vertical resolution of the GPS occultation profiles along with its cloud-penetrating ability could potentially be used to provide insights into the boundary layer processes.

Summary

The GPS occultation technique offers unique capabilities in the remote sensing of the Earth's atmosphere. It provides high vertical resolution, all-weather profiling data from the planetary boundary layer to the stratosphere that are valuable for a wide range of weather and climate

applications. In particular, the measurements possess qualities that make them particularly useful for monitoring long-term climate change. The future of GPS occultation looks exciting, with a proliferation of GNSS transmitters from the European Galileo, Russian GLONASS, and Chinese Compass navigation satellites. Increased spatial-temporal coverage would especially benefit studies of tropical storms and other mesoscale phenomena.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Anthes, R. A., et al., 2008. The COSMIC/FORMOSAT-3 mission: early results. *Bulletin of the American Meteorological Society*, **89**, 313–333.
- Ao, C. O., Hajj, G. A., Iijima, B. A., Mannucci, A. J., Schröder, T. M., de la Torre Juárez, M., and Leroy, S. S., 2006. Sensitivity of stratospheric retrievals from radio occultations on upper boundary conditions. In Foelsche, U., Kirchengast, G., and Steiner, A. (eds.), *Atmosphere and Climate: Studies by Occultation Methods*. Berlin: Springer, pp. 17–26.
- Ao, C. O., Hajj, G. A., Meehan, T. K., Dong, D., Iijima, B. A., and Mannucci, A. J., 2009. Rising and setting GPS occultations by use of open-loop tracking. *Journal of Geophysical Research*, **114**, D04101, doi:10.1029/2008JD010483.
- Ao, C. O., Waliser, D. E., Chan, S. K., Li, J.-L., Tian, B., Xie, F., and Mannucci, A. J., 2009. Planetary boundary layer heights from GPS radio occultation refractivity and humidity profiles. *Journal of Geophysical Research*, **117**, D16117, doi:10.1029/2012JD017598, 2012.
- Basha, G., and Ratnam, M. V., 2009. Identification of atmospheric boundary layer height over a tropical station using high resolution radiosonde refractivity profiles: comparison with GPS radio occultation measurements. *Journal of Geophysical Research*, **114**, D16101, doi:10.1029/2008JD011692.
- Cucurull, L., Derber, J. C., Treadon, R., and Purser, R. J., 2008. Preliminary impact studies using Global Positioning System radio occultation profiles at NCEP. *Monthly Weather Review*, **136**, 1865–1877.
- Fjeldbo, G., Kliore, A. J., and Eshleman, V. R., 1971. The neutral atmosphere of Venus as studied with the Mariner V radio occultation experiments. *Astronomy Journal*, **76**, 123–140.
- Goody, R., Anderson, J., and North, G., 1998. Testing climate models: an approach. *Bulletin of the American Meteorological Society*, **79**(11), 2541–2549.
- Gorbunov, M. E., 2002. Canonical transform method for processing radio occultation data in the lower troposphere. *Radio Science*, **37**, 5, doi:10.1029/2000RS002592.
- Guo, P., Kuo, Y.-H., Sokolovskiy, S. V., and Lenschow, D. H., 2011. Estimating atmospheric boundary layer depth using COSMIC radio occultation data. *Journal of Atmospheric Science*, **68**, 1703–1713.
- Gurvich, A. S., and Krasil'nikova, T. G., 1987. Navigation satellites for radio sensing of the earth's atmosphere (in Russian). *Soviet Journal of Remote Sensing*, **6**, 89–93.
- Hajj, G. A., Kursinski, E. R., Romans, L. J., Bertiger, W. I., and Leroy, S. S., 2002. A technical description of atmospheric sounding by GPS occultation. *Journal of Atmospheric and Solar-Terrestrial Physics*, **64**(4), 451–469.

- Hajj, G. A., Ao, C. O., Iijima, B. A., Kuang, D., Kursinski, E. R., Mannucci, A. J., Meehan, T. K., Romans, L. J., de la Torre Juárez, M., and Yunck, T. P., 2004. CHAMP and SAC-C atmospheric occultation results and intercomparisons. *Journal of Geophysical Research*, **109**, D6, doi:10.1029/2003JD003909.
- Healy, S. B., and Eyre, J. R., 2000. Retrieving temperature, water vapour and surface pressure information from refractive-index profiles derived by radio occultation: a simulation study. *Quarterly Journal of the Royal Meteorological Society*, **126**(566), 1661–1683.
- Healy, S. B., and Thépaut, J. N., 2006. Forecast impact experiment with GPS radio occultation measurements. *Quarterly Journal of the Royal Meteorological Society*, **132**, 605–623.
- Jensen, A. S., Lohmann, M. S., Benzon, H.-H., and Nielsen, A. S., 2003. Full spectrum inversion of radio occultation signals. *Radio Science*, **38**, 3, doi:10.1029/2002RS002763.
- Kuo, Y.-H., Wee, T., Sokolovskiy, S., Rocken, C., Schreiner, W., Hunt, D., and Anthes, R., 2004. Inversion and error estimation of GPS radio occultation data. *Journal of the Meteorological Society of Japan*, **82**(1B), 507–531.
- Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P., and Hardy, K. R., 1997. Observing earth's atmosphere with radio occultation measurements using the Global Positioning System. *Journal of Geophysical Research*, **102**(D19), 23429–23465.
- Nishida, M., Shimizu, A., Tsuda, T., Rocken, C., and Ware, R. H., 2000. Seasonal and longitudinal variations in the tropical tropopause observed with the GPS occultation technique (GPS/MET). *Journal of the Meteorological Society Japan*, **78**, 691–700.
- Poli, P., Healy, S. B., and Dee, D. P., 2010. Assimilation of Global Positioning System radio occultation data in the ECMWF ERA–Interim reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1972–1990, doi:10.1002/qj.722.
- Randel, W. J., Wu, F., and Rios, W. R., 2003. Thermal variability of the tropical tropopause region derived from GPS/MET observations. *Journal of Geophysical Research*, **108**(D1), 4024.
- Schmidt, T., Wickert, J., Beyerle, G., and Reigber, C., 2004. Tropical tropopause parameters derived from GPS radio occultation measurements with CHAMP. *Journal of Geophysical Research*, **109**, D13105.
- Schreiner, W., Rocken, C., Sokolovskiy, S., Syndergaard, S., and Hunt, D., 2007. Estimates of the precision of GPS radio occultations from COSMIC/FORMOSAT-3. *Geophysical Research Letters*, **34**, 4, doi:10.1029/2006GL027557.
- Schröder, T., Leroy, S., Stendel, M., and Kaas, E., 2003. Validating the microwave sounding unit stratospheric record using GPS occultation. *Geophysical Research Letters*, **30**(14), 1734, doi:10.1029/2003GL017588.
- Smith, E. K., and Weintraub, S., 1953. The constants in the equation for atmospheric refractive index at radio frequencies. *Proceedings of the IRE*, **41**, 1035–1037.
- Sokolovskiy, S., 2001. Modeling and inverting radio occultation signals in the moist troposphere. *Radio Science*, **36**, 3, doi:10.1029/1999RS002273.
- Sokolovskiy, S., Kuo, Y.-H., Rocken, C., Schreiner, W. S., Hunt, D., and Anthes, R. A., 2006. Monitoring the atmospheric boundary layer by GPS radio occultation signals recorded in open-loop mode. *Geophysical Research Letters*, **33**, L12813, doi:10.1029/2006GL025955.
- Son, S.-W., Tandon, N. F., and Polvani, N. F., 2011. The fine-scale structure of the global tropopause derived from COSMIC GPS radio occultation measurements. *Journal of Geophysical Research*, **116**, D20, doi:10.1029/2011JD016030.
- Steiner, A. K., Kirchengast, G., Borsche, M., Foelsche, U., and Schoengassner, T., 2007. A multi-year comparison of lower stratospheric temperatures from CHAMP radio occultation data with MSU/AMSU records. *Journal of Geophysical Research*, **112**, D22110, doi:10.1029/2006JD008283.
- von Engel, A., Healy, S., Marquardt, C., Andres, Y., and Sancho, F., 2009. Validation of operational GRAS radio occultation data. *Geophysical Research Letters*, **36**, 17, doi:10.1029/2009GL039968.
- Vorob'ev, V. V., and Krasil'nikova, T. G., 1994. Estimation of the accuracy of the atmospheric refractive index recovery from Doppler shift measurements at frequencies used in the NAVSTAR system. *Physics of the Atmosphere and Ocean*, **29**, 602–609.
- Ware, R., et al., 1996. GPS sounding of the atmosphere from low earth orbit – preliminary results. *Bulletin of the American Meteorological Society*, **77**, 19–40.
- Wickert, J., Reigber, C., Beyerle, G., König, R., Marquardt, C., Schmidt, T., Grunwaldt, L., Galas, R., Meehan, T. K., Melbourne, W. G., and Hocke, K., 2001. Atmospheric sounding by GPS radio occultation: first results from CHAMP. *Geophysical Research Letters*, **28**(17), 3263–3266.
- Yunck, T. P., Lindal, G. F., and Liu, C. H., 1988. The role of GPS in precise earth observation. In *IEEE Position, Location and Navigation Symposium*, November 29–December 2, Orlando.

ICE SHEETS AND ICE VOLUME

Robert Thomas
Sigma Space, Gorzow Wlkp, Poland

Definition and introduction

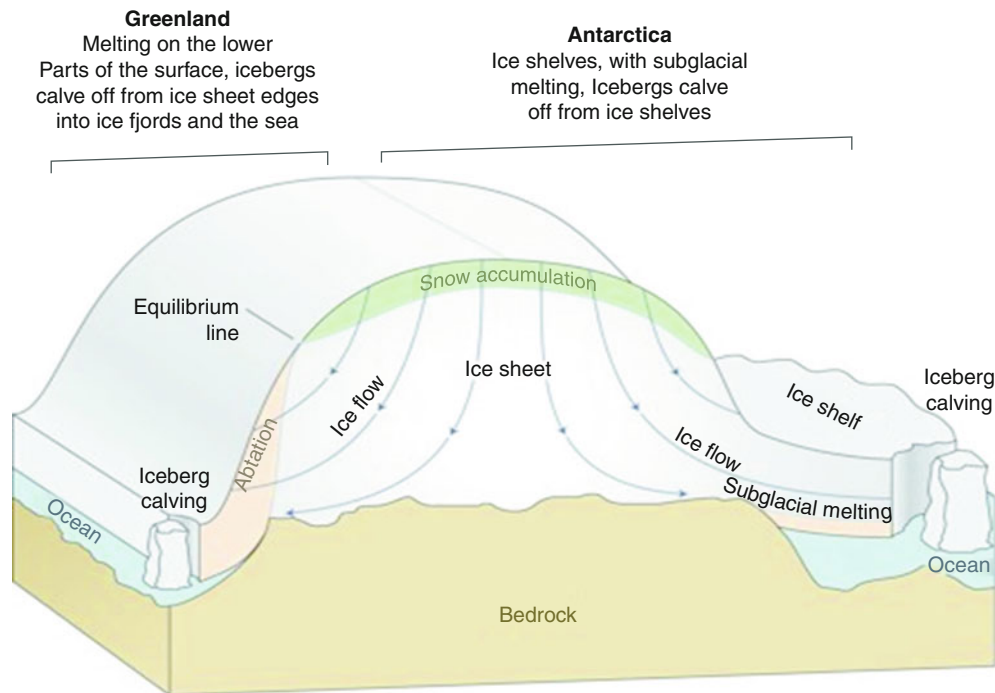
An *ice sheet* is a large mass of ice resting on land that is continental or subcontinental in extent, with the ice thick enough to cover most of the underlying bedrock topography. Its shape is mainly determined by the dynamics of its gravity-driven outward flow. There are only two ice sheets in the modern world, in Greenland and Antarctica, but during glacial periods there were others. At the beginning of the *Cenozoic*, some 65 million years (Ma) ago, neither Greenland nor Antarctica supported an ice sheet. Then, as greenhouse gases in the atmosphere decreased, slow, irregular cooling allowed ice masses to accumulate and survive, first as mountain glaciers and then with the first continental ice sheet forming over Antarctica as early as 33 Ma ago. Further cooling led to extensive ice formation on Arctic land areas about 2.6 Ma ago, initiating the *Quaternary* series of ice ages with warmer interglacials at roughly 0.12 Ma intervals. Cycling between cold *glacial* and warmer *interglacial* periods is driven by periodic features of Earth's orbit, with ice sheets growing when sunshine shifts away from the Northern Hemisphere and melting when northern sunshine returns. These changes are amplified by feedbacks, such as greenhouse gas concentrations that rise and fall as the ice cover shrinks and grows, and greater reflection of sunshine caused by more extensive ice ([Figure 1](#)).

Human civilization developed during the most recent (*Holocene*) interglacial, extending over the past 11,000–12,000 years (ka). Holocene warming was interrupted during the Little Ice Age (about 1250–1850 AD), when the Greenland ice sheet and most Arctic glaciers reached

their maximum Holocene extent. Since then, warming has resulted in Arctic-wide glacier recession, with similar responses becoming apparent more recently in Antarctica. These trends will continue if greenhouse gas concentrations continue to increase into the future ([Table 1](#)).

During the previous interglacial, 130–120 ka ago, Arctic summers were about 5 °C warmer than now and the Greenland ice sheet was considerably smaller than at present, with probable losses also from Antarctica, resulting in global sea level some 4–6 m above present values.

An ice sheet forms by the continual accumulation of snow on its surface. Over parts of an ice sheet, summers are warm enough to melt some of this snow, with some of the melt water percolating into the snow and refreezing and some running off the ice sheet into the ocean. The difference between local precipitation and melt-water runoff is the *surface mass balance*. As successive layers of snow build up, the layers beneath are gradually compressed into solid ice, at depths up to several tens of meters. The snow input is approximately balanced by glacial outflow, so the height of the ice sheet stays roughly constant through time. The difference between the mass of ice added by snowfall and that lost by runoff and ice motion is the *total mass balance*. The ice is driven downhill by gravity, from the highest points in the interior to the coast, where it melts at lower, warmer elevations or breaks off as icebergs. In contrast to Greenland, losses by surface melting are small in Antarctica and are primarily from the northern Antarctic Peninsula and the northernmost fringes of East Antarctica. The ice sheet spreads under its own weight, partly by *internal deformation* that increases with the cube of the driving stress which is proportional to the product of ice thickness and ice-surface slope. Internal deformation predominates where the ice is frozen to its bed and is typically slow. Where the bed of the ice sheet is at the melting point, ice may move more rapidly by



Ice Sheets and Ice Volume, Figure 1 Ice sheets showing the main features of those in Greenland (left of Figure) and Antarctica (right) (Source: K. Steffen, CIRES/University of Colorado).

Ice Sheets and Ice Volume, Table 1 The present areas and volumes of the two ice sheets, with estimates of the potential sea-level rise if they were to melt completely

	Area covered (million km ²)	Ice volume (million km ³)	Potential sea-level rise(m)
Ice sheets (total)	14.0	28	64
Greenland	1.7	3	7
Antarctica	12.3	25	57

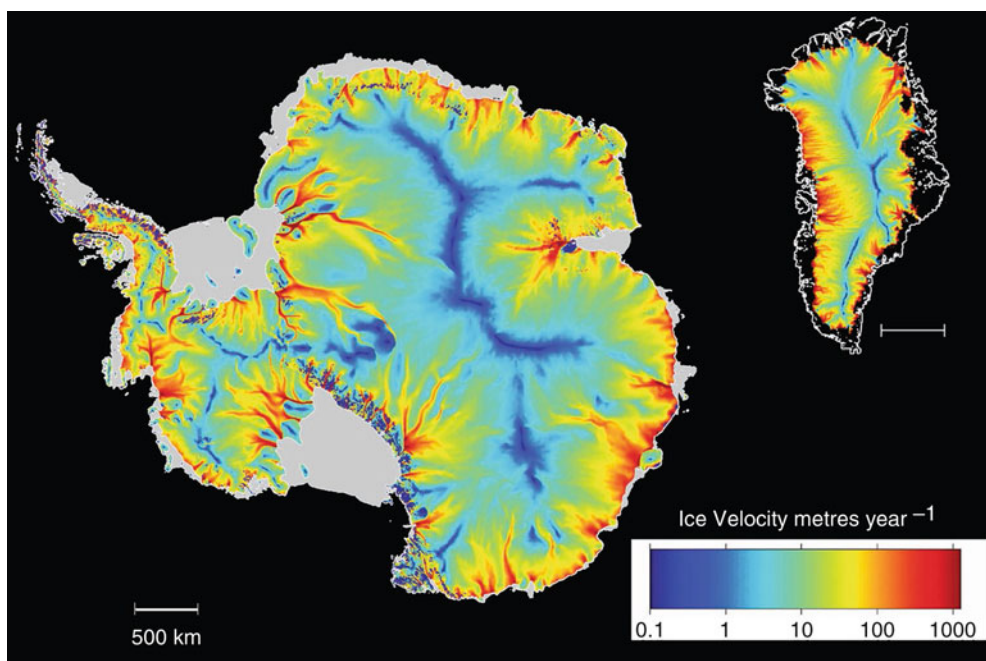
sliding over its bed or by the deformation of water-saturated basal moraine and mud. This rapidly moving ice is generally channeled along bedrock troughs, which become further deepened as they are eroded by the moving ice. *Ice streams* are “jets” of faster ice flanked by slower flowing parts of an ice sheet, becoming outlet glaciers near the coast, where many are flanked by rock (Figure 2). Thus, for example, the fast-moving Jakobshavn Isbrae starts as an ice stream about 100 km inland that flows along a deep bedrock trough but is flanked by rock before calving *icebergs* into a long fjord on the west coast of Greenland.

The Antarctic ice sheet

Antarctica is the fifth largest and most southerly continent and is generally divided into two major regions: *West Antarctica*, including the *Antarctic Peninsula*, in the Western

Hemisphere and *East Antarctica* in the Eastern Hemisphere. The *Transantarctic Mountains* separate the two, extending 3,000 km between the Atlantic and Pacific coasts. Most of Antarctica is covered by ice, with thick, grounded *inland ice* resting on a solid bed, and the floating *ice shelves* formed partly from ice flowing onto the ocean across the *grounding line* and partly from snow that falls onto them. The inland ice and the ice shelves together constitute the ice sheet, which is surrounded by ephemeral *sea ice*. Inland ice (ice shelf) ranges in thickness up to 5 km (1.5 km), with an average of about 2,400 m (600 m). Antarctica is, on average, by far the highest of the continents, with surface elevations rising to 4,000 m in central East Antarctica. The South Pole lies within the continent, which is completely surrounded by the ocean. The ice sheet was considerably larger during the Last Glacial Maximum, some 20,000 years ago, and retreated to near its present extent within the last several thousand years and perhaps is still retreating (Figure 3).

Separation of Southern Hemisphere continents from *Gondwanaland* finally left Antarctica isolated when *Drake's Passage* separated South America from the Antarctic Peninsula about 30 million years ago. This allowed uninterrupted circulation of the *Circumpolar Current* in the *Southern Ocean*. The combination of continuous oceanic and atmospheric circumpolar circulation around Antarctica led to cooling of the now isolated continent and formation of the ice sheet. Outflow from the inland ice is organized into a series of drainage basins, separated by



Ice Sheets and Ice Volume, Figure 2 Estimated “balance” ice velocities for the ice sheets in Antarctica (*left*) and Greenland (*right*). Note that the surface velocity scale is logarithmic. The *gray* areas around Antarctica are floating ice shelves. The *black* coastal areas around Greenland are ice-sheet-free mountains, which include some glaciers and ice caps. Ice discharge at these velocities exactly balances the mass added by snowfall (Allison et al., 2009, copyright Antarctic Science Ltd 2009).



Ice Sheets and Ice Volume, Figure 3 The Antarctic ice sheet and ice shelves. Some marginal seas are named, and Drake’s Passage is to the north of the Antarctic Peninsula (Bentley et al., 2007).

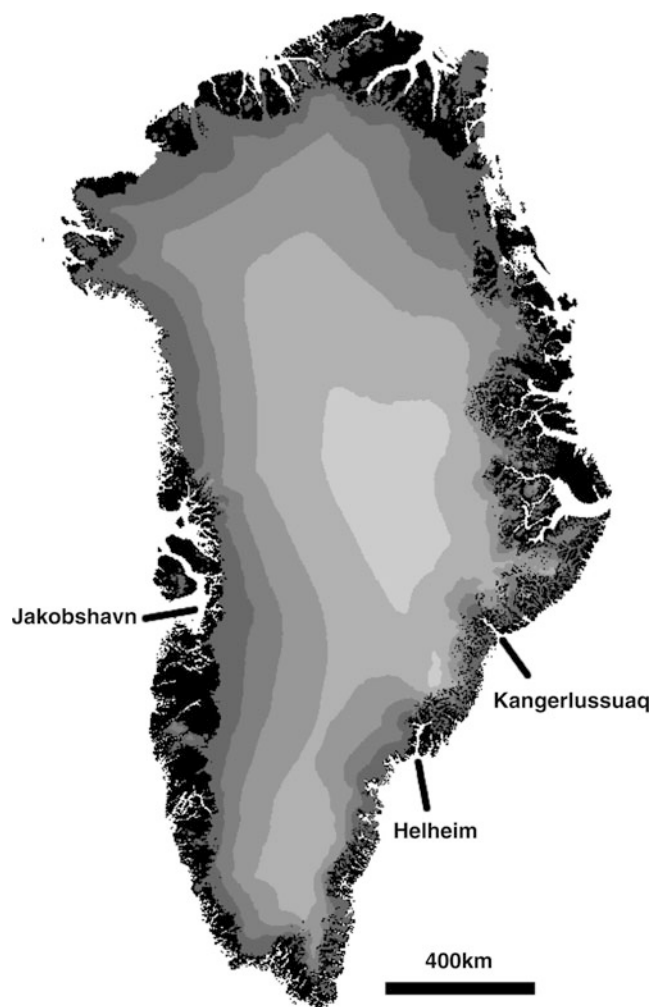
ice divides, much like water flow on other continents. Most drainage basins concentrate the flow into either narrow outlet glaciers, particularly through the Transantarctic Mountains, or fast-moving ice streams, which are surrounded by slow moving ice rather than rock walls. Flow speeds increase toward the coast, because of increasing total upstream snowfall and because the ice generally thins to seaward. Where there is convergence into an outlet glacier or ice stream, the speed of flow is increased even more and can reach more than a kilometer per year.

The *West Antarctica ice sheet* (WAIS) flows mostly into the *Ross Ice Shelf* in the Ross Sea, the *Filchner/Ronne Ice Shelf* in the Weddell Sea, and smaller ice shelves in the *Amundsen Sea* which lies between the Bellingshausen and Ross Seas. The two big ice shelves are each about the area of Spain. Driven by gravity, ice shelves spread out over the ocean under their own weight, supplied with ice partly from tributary glaciers and ice streams, partly from snow that falls on their surfaces, and in some cases partly from ice frozen at basal interfaces with the ocean. Basal melting is far more common than freezing and can be as high as tens of meters per year and accounts for a large fraction of total losses, with iceberg calving from seaward ice fronts responsible for the rest. Where ice shelves run aground on shoaling seabed, *ice rumpled* or *ice rises* are formed. Ice rumpled are dragged along by surrounding ice shelf to move slowly over the sea bed; ice rises are stationary and dome shaped, with their own patterns of outward ice motion.

The Greenland ice sheet

The Greenland ice sheet occupies a latitude band extending from 60 to 80 N and covers an area of 1.7 million square km. With an average thickness of 1,600 m, it has a total volume of approximately 3 million cubic km – equivalent to a sea-level rise of about 7 m. It consists of a northern dome and a southern dome, with maximum surface elevations of approximately 3,200 m and 2,850 m, respectively, linked by a long saddle with elevations around 2,500 m. Bedrock beneath much of the ice sheet is close to sea level, but the ice sheet is fringed almost completely by coastal mountains through which the ice sheet is drained by many glaciers (Figure 4).

The ice sheet in Greenland differs substantially from that in Antarctica, which is almost ten times larger in volume. Antarctica straddles the South Pole and has a dominant influence on its own climate and on the surrounding ocean, with cold conditions even during the summer and around its northern margins. Away from the coast, much of Antarctica is a cold desert, with very low precipitation rates. There is little surface melting, even near the coast, and most of the melt water soaks into underlying snow and refreezes. Because of the cold conditions, vast floating ice shelves exist around much of the continent. Ice drainage is primarily by glaciers and ice streams, some of which penetrate deep into the heart of the ice sheet, moving at maximum speeds of a few



Ice Sheets and Ice Volume, Figure 4 The Greenland ice sheet, showing key glaciers. Black denotes rock, and gray denotes ice, shaded in 500 m elevation bands with the highest contour at 3,000 m (Prepared by S. Manizade).

hundred meters per year, apart from a few exceptions. Most glaciers and ice streams flow into ice shelves, which are also fed by snow accumulation on their surfaces. The ice shelves thin toward their seaward ice fronts, partly by ice creep and partly by basal melting, with melting rates generally of a few tens of centimeter/year, but increasing to tens of meters/year near the grounding lines of some ice shelves. Thus, ice loss from the Antarctic ice sheet is primarily by melting from beneath, and iceberg calving from seaward ice fronts, of ice shelves.

By contrast, the Greenland climate is strongly affected by its proximity to other land masses and to the North Atlantic, with the Gulf Stream to the south and regions of North Atlantic deep water production to the east and west. Ice-core data from the summit of the ice sheet indicate that Greenland temperatures and accumulation rates can increase significantly over periods of a few years

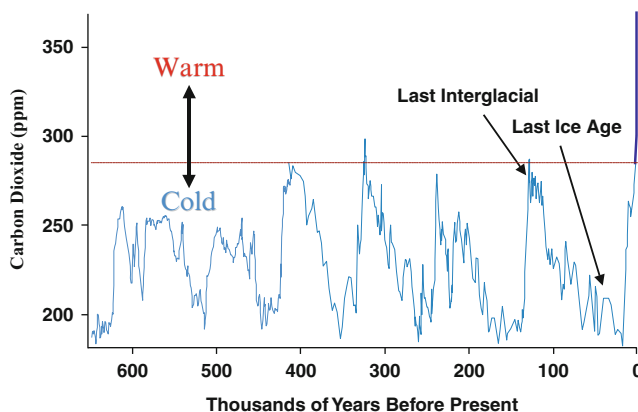
to decades (Alley et al., 1993). Other major contrasts with Antarctica include widespread summer melting, higher accumulation rates, very few ice shelves, and faster, narrower glaciers. Summer melting occurs over about 50 % of the ice-sheet surface, depending on summer temperatures, with much of the resulting melt water flowing into the sea, either along channels cut into the ice surface or by draining to the bed via crevasses. Mass loss by surface melting accounts for about half the total ice loss from Greenland, with iceberg calving responsible for the rest. Most Greenland outlet glaciers are narrower, by an order of magnitude, than their Antarctic counterparts, but some reach speeds that are an order of magnitude higher. Consequently, these glaciers drain very large volumes of ice, with discharge rates strongly determined by fast-glacier dynamics, which are poorly understood.

The polar ice sheets are remote, difficult to visit, and easily regarded as being of little significance to the rest of the world. But in two ways, they have turned out to be of growing importance: They preserve a record of past climate and climate change, and they control global sea level.

Ice cores reveal a history of climate

Most of the surface of the Antarctic ice sheet, and about half of the Greenland ice sheet, is too cold for there to be much surface melting, and snow that falls on the ice sheets is slowly compressed until air that is trapped in the snow becomes isolated into small bubbles that shrink in size at increasing depths and finally dissolve in surrounding ice. Because there is little or no surface melting, annual layers of snow accumulation are unmixed with others, and they retain a signature of past climate, which can be sampled by drilling deep ice cores from cold parts of the ice sheet. The variety of climatic proxies in these cores is greater than in any other natural recorder of climate, such as tree rings or sediment layers. They include information on past ice-sheet surface temperatures and elevations, precipitation, chemistry and gas composition of the lower atmosphere, volcanic eruptions, solar variability, sea-surface productivity, desert extent, and forest fires.

Prevailing temperatures when the snow fell can be inferred from the ratios of different isotopes of oxygen and hydrogen in its water molecules. The ocean contains both normal and “heavy” water: Roughly one molecule in 500 includes at least one extra neutron in the nucleus of an oxygen or hydrogen atom. The lighter molecules evaporate more easily, and the heavier molecules condense more easily. Consequently, as water that evaporated from the ocean is carried inland over an ice sheet, the heavy molecules preferentially rain or snow out, leaving the lighter molecules to precipitate at colder temperatures. Calibration of the ratios of heavy versus light molecules in present-day snowfall against prevailing air temperatures allows interpretation of similar ratios in snow and ice recovered by the ice cores. Because of substantial



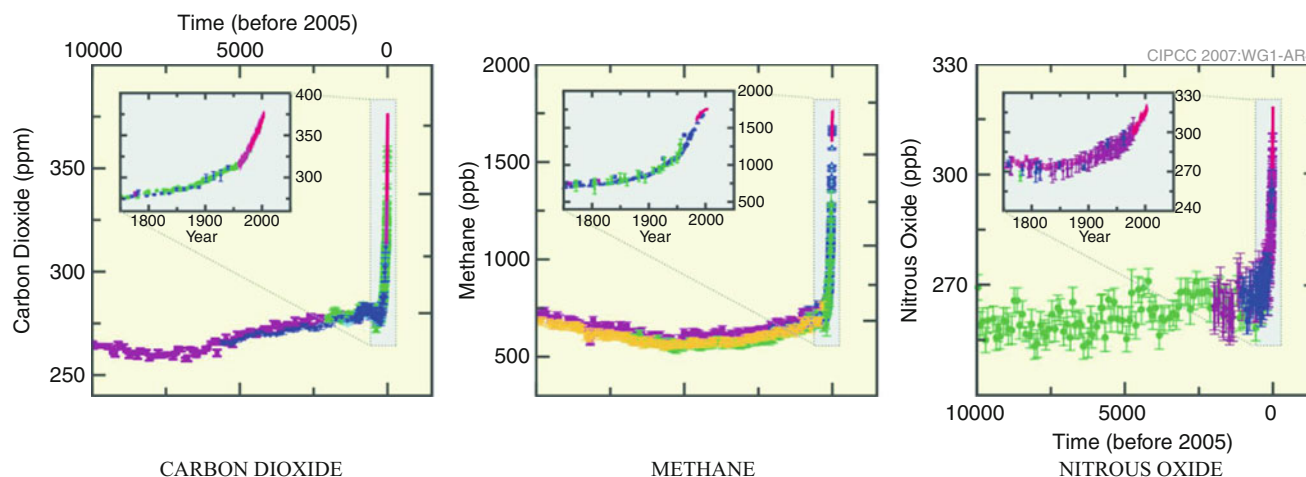
Ice Sheets and Ice Volume, Figure 5 Ice ages, represented by variations in atmospheric carbon dioxide preserved in ice cores, are “forced” by Earth’s orbital clock, which changes the patterns of received sunlight. The big increase in carbon-dioxide concentrations over the past 100 years is shown by the dark blue vertical line.

differences between temperatures of summer and winter snowfall, annual layers can be counted down to considerable depths in regions of high snowfall, such as Greenland.

In the upper parts of a core, annual layers may also be visible, related to dustiness or summer-melt intensity. But deeper layers thin because of outward spreading of the ice sheet, and eventually individual years cannot be distinguished. Dating can also be inferred from the depths of events such as nuclear bomb fallout in the upper levels and ash layers corresponding to known volcanic eruptions. At greater depths, approximate dating can be estimated from radiocarbon or other radionuclide dating or by ice-flow modeling.

In addition to information from the isotope composition of the ice, the air bubbles and dust trapped in the ice cores allow for measurement of the atmospheric concentrations of dust and trace gases, including the greenhouse gases carbon dioxide, methane, and nitrous oxide. Total atmospheric concentration within the core at different depths indicates the air pressure, and hence surface elevation, at the time of snow deposition. Analysis of the dust and its concentration provides information on its source, and hence wind patterns, and on overall global dustiness. It also gives a strong indication of human influence on the concentration of atmospheric pollutants, such as lead and DDT. Greenhouse gas concentrations clearly show growth in these gases since the beginning of the Industrial Revolution, with a progressive acceleration in this growth as a burgeoning world population demands the fruits of affluence. Greenhouse gas concentrations are now higher than at any time during the 800,000 years covered by the ice-core record (Figures 5 and 6).

The length of the time record obtained by an ice core depends on the depth reached by the core and by the local snow-accumulation rate. It varies from a few years for



Ice Sheets and Ice Volume, Figure 6 Ice cores show these dramatic increases in greenhouse gases over the past 100 years as global energy consumption has increased.

a shallow core (typically 10 m) up to 800,000 year for the 3,200 m EPICA core in East Antarctica, which provides a climate record covering eight glacial cycles. The transition from glacial to interglacial conditions about 430,000 years ago resembles the transition into the present interglacial period in terms of the magnitude of change in temperatures and greenhouse gases. The interglacial that followed lasted 28,000 years compared to the 12,000 years recorded so far in the present interglacial period. Given the similarities between this earlier warm period and today, this may imply that, without human intervention, a climate similar to the present one would extend well into the future.

Although ice cores have revealed much about the growth and collapse of past ice sheets, little was known about their present behavior until quite recently. The past two decades have seen a dramatic change in our ability to observe the ice sheets, made possible by advances in remote-sensing techniques and their application aboard aircraft and satellites.

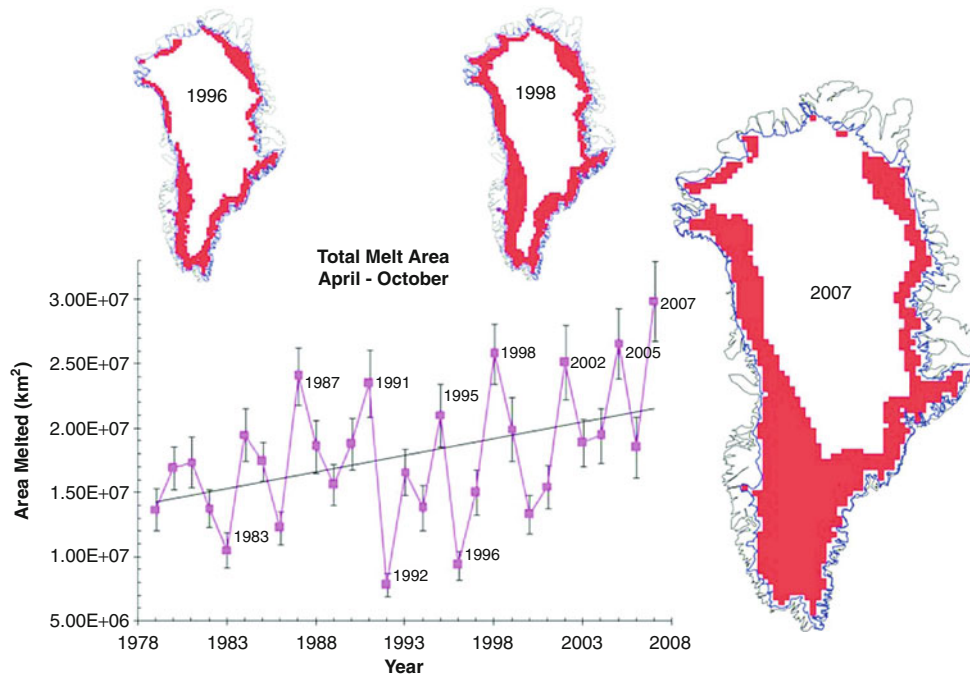
Applications of remote sensing to ice-sheet investigations

The polar ice sheets are remote and inhospitable and, until the 1970s, observations were largely made in situ, supplemented by just a few air photographs. Consequently, little was known about their behavior apart from the very few locations where detailed measurements had been made. Snow accumulation was measured by repeated stake measurements or by interpreting snow layers in a pit or shallow drill core. Ice motion was measured by repeatedly surveying stake locations with respect to local fixed points, by sun or star observations where there were no fixed points, or by repeated mapping of local magnetic anomalies. Ice thickness was measured

using seismic techniques. Progress was slow until, in the early 1970s, scientists at the Scott Polar Research Institute in Cambridge, England, developed a low-frequency airborne radar capable of penetrating ice and receiving a reflected signal from the ice bed. This made possible major ice-thickness surveys in both Antarctica and Greenland that advanced our knowledge substantially. Later in the same decade, NASA launched *SeaSat*, a near-polar orbiting satellite with various sensors aboard designed to observe the ocean but which also turned out to be remarkably well suited to observing the polar ice sheets.

Included aboard *SeaSat* were:

- A *passive-microwave* radiometer, which images microwave emissions from the surface. Passive-microwave data have been successfully applied to the mapping of surface melting on the ice sheets and can also be used to help interpolate estimates of snow-accumulation rates between in situ measurements (Figure 7).
- A *scatterometer*, which images microwave reflections from an onboard radar and maps areas with near-surface ice layers in addition to those with surface melting.
- A *synthetic aperture radar* (SAR), which records the time delay, the phase, and the amplitude of reflections from a side-looking radar. In addition to providing high-resolution, all-weather images of glaciers and ice sheets, the information from repeated SAR images of the same area can be used to reconstruct surface topography and to map accurate estimates of ice motion at high spatial resolution. This final capability transformed our ability to study the polar ice sheets.
- A *radar altimeter* that measures the closest range from the spacecraft to the surface beneath. Designed to measure ocean topography, this instrument was the first to provide estimates of ice thickening/thinning rates over



Ice Sheets and Ice Volume, Figure 7 The graph shows the total cumulative melt area from 1979 to 2007 for the Greenland ice sheet derived from satellite passive-microwave data. The map inserts display the area of melt for 1996, 1998, and the record year 2007 (From K. Steffen, CIRES, University of Colorado).

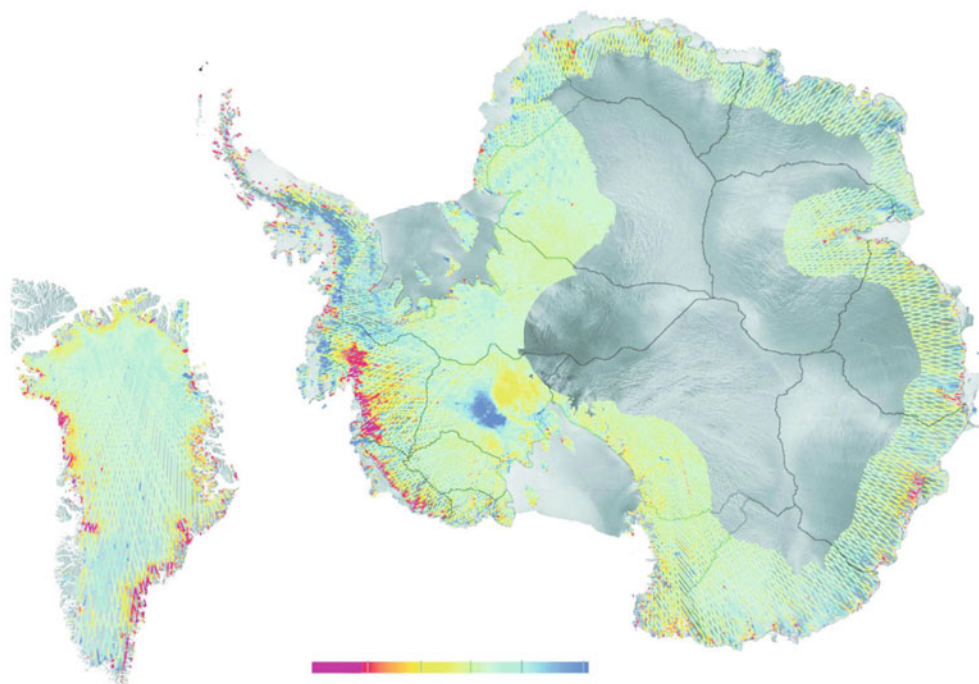
large regions. Unfortunately, however, the large radar beam width and the possibility of radar penetration below the ice-sheet surface make accurate interpretation very difficult.

All of these sensors clearly demonstrated the enormous potential of satellite remote sensing for ice-sheet research and have been included on various other spacecraft since. Together, they have revolutionized the study of the polar ice sheets. They have also helped focus in situ measurements by identifying regions where such measurements may be most productive. Nevertheless, some measurements are best made from aircraft, particularly after the availability for civil applications of the *Geodetic Positioning System* (GPS) in the early 1990s made it possible to reconstruct very accurately an aircraft trajectory and to follow earlier aircraft routes during repeat surveys. This capability made it possible, starting in 1991, for NASA to make accurate measurements of ice-surface elevations (to within ± 10 cm), using a scanning laser altimeter along several tens of thousands of km of flight tracks over the Greenland ice sheet. Repeat surveys in 1993 and 1998 provided the first indication that the ice sheet was losing mass overall, with slow central thickening more than balanced by far more rapid thinning nearer the coast. Ice thickness was also measured along these flight lines, providing information needed to interpret

observed thickening/thinning rates, information that cannot yet be acquired from satellite-borne sensors. Unfortunately, the remoteness and sheer size of Antarctica prevented similar surveys over the Antarctic ice sheet.

Later, in 2003, NASA launched the Geoscience Laser Altimeter System (GLAS) aboard the near-polar orbiting *ICESat* to make similar measurements over both ice sheets and the rest of Earth's surface. Unfortunately, problems with the laser system resulted in reducing its use to two or three surveys per year, each lasting about a month. The laser finally ceased operating in 2009. Data from *ICESat* confirmed the overall pattern of slow thickening over parts of central Greenland and rapid thinning over many coastal regions and revealed patterns of elevation change over Antarctica. However, cloud cover and the quite large separation between adjacent *ICESat* orbits resulted in incomplete coverage of the ice sheets, particularly in coastal regions where thinning rates are highest (Figure 8).

A joint NASA/German Aerospace Center (DLR) mission, the Gravity Recovery and Climate Experiment (*GRACE*), launched in 2002 and expected to continue operating until 2015, maps Earth's gravity field by precisely measuring the time-varying distance between two small satellites chasing each other along a near-polar orbit. Temporal changes in the gravity field result from mass shifting over, and beneath, the Earth's surface, including



Ice Sheets and Ice Volume, Figure 8 Rate of change of surface elevation for Antarctica and Greenland; the color scale shows rates of elevation change ranging from thinning of 1.5 m/year (purple) to thickening of 0.5 m/year (blue) (Pritchard et al., 2009).

the changing mass, of the polar ice sheets. Results have been widely used to infer changes in ice mass or total mass balance integrated over very large areas.

Until quite recently, the ice sheets were assumed to be approximately in *steady state*, with total snow accumulation roughly in balance with total losses by melting and ice discharge. This was not based on measurements of these parameters, but rather on very low rates of observed sea-level rise (SLR) – close to only 1 mm/year – implying very small changes in ice-sheet volume. But measuring and understanding the mass balance of the Greenland and Antarctic ice sheets and of individual drainage basins within the ice sheets have long been major goals of ice-sheet science, and the satellite and aircraft measurements described above resulted in major advances toward these goals. Fortuitously, these advances came at a time when SLR rates began to increase substantially. If the polar ice sheets were to melt completely, global sea level would rise by about 64 m, so measuring and understanding their mass balance help to explain the recent increase in SLR and to predict future SLR.

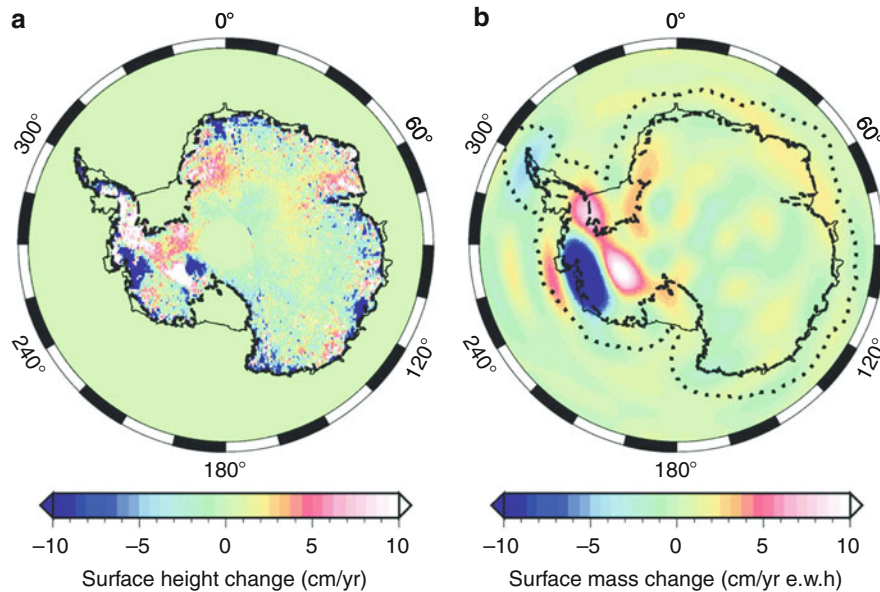
Measuring the mass balance of the polar ice sheets

Techniques for measuring total mass balance of the polar ice sheets include:

- *The mass-budget approach*, comparing gains by surface and internal accumulation with losses by ice discharge, sublimation, and melt-water runoff. Snow accumulation is estimated from stake measurements,

annual layering in ice cores, or from regional atmospheric climate modeling. Ice discharge is the product of velocity and thickness, with velocities measured by ground-based survey, photogrammetry, or, increasingly, with satellite SAR operating interferometrically, and thickness generally measured by airborne radar. Melt-water runoff is inferred from regional atmospheric climate models validated with surface observations where available. Mass-budget calculations involve the comparison of two very large numbers, and small errors in either can result in large errors in estimated total mass balance.

- *Repeated altimetry surveys*, to measure height changes, from which changes in volume and mass are inferred. Rates of surface-elevation change with time (dS/dt) reveal changes in ice-sheet mass after correction for changes in depth/density profiles and vertical bedrock motion. Satellite radar altimetry has been widely used, but data interpretation is difficult because of time-variable radar penetration into the snow and, over sloping and undulating surfaces, because of the large radar footprint. These limitations are overcome by laser altimetry, but problems with NASA's ICESat resulted in a comparatively sparse data set that poorly reveals the behavior of individual glaciers and ice streams, which are better investigated using airborne laser altimetry, with the added advantage of providing an opportunity to measure ice thickness at the same time. All altimetry mass-balance estimates include substantial uncertainties associated with settling of surface snow



Ice Sheets and Ice Volume, Figure 9 Estimates of rates of surface-elevation changes from ICESat (a), and mass changes from GRACE represented as equivalent water depth changes (b) (Riva et al., 2009).

and with density assumed to convert thickness changes to mass changes.

- *Satellite measurements of temporal changes in gravity*, to infer mass changes directly. Since 2002, the GRACE satellite has measured Earth's gravity field and its temporal variability. After removing the effects of tides, atmospheric loading, spatial and temporal changes in ocean mass, etc., high-latitude data contain information on temporal changes in the mass distribution of the ice sheets and underlying rock. Mass-balance estimates are – at coarse resolution – several hundred kilometers, but this has the advantage of covering entire ice sheets, which is extremely difficult using other techniques. Consequently, GRACE estimates include mass changes on the many small ice caps and isolated glaciers that surround the big ice sheets, which are strongly affected by changes in the coastal climate. Error sources include measurement uncertainty, leakage of gravity signal from regions surrounding the ice sheets, and causes of gravity changes other than ice-sheet changes. Of these, the most serious are the gravity changes associated with vertical bedrock motion (Figure 9).

Mass balance of the Greenland and Antarctic ice sheets

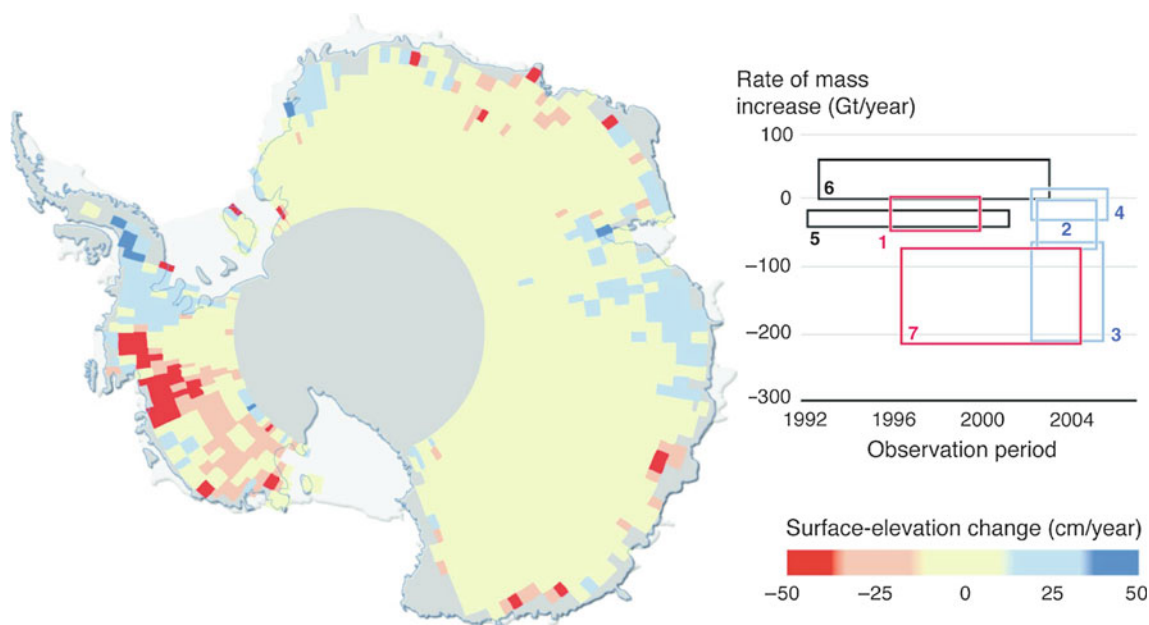
Most model results previously suggested that climate warming would result primarily in increased melting from coastal regions and an overall increase in snowfall, with probably a small mass loss from Greenland and a small gain in Antarctica during the twenty-first century, and

little combined impact on sea level. But recent measurements using the techniques described in the last section show substantial losses from both ice sheets but with poor agreement between estimates from the three different techniques. Moreover, each technique is under continual development and improvement, so that estimates based on the same set of measurements are continually changing. Nevertheless, the three independent techniques for estimating mass balance do provide the means for cross-checking results, and they increase our confidence in results where there is general agreement.

Available results indicate that annual ice loss from the Greenland ice sheet during the period with good observations increased from a few tens of cu km in the mid-1990s to perhaps as much as 200 cu km for the most recent observations. Near-coastal summer melting has increased substantially, but this is largely balanced by increased snowfall further inland, and an increasing proportion of the ice loss is by enhanced ice discharge down accelerating glaciers. Over the same period, net loss from Antarctica increased from probably near zero in the mid-1990s to nearly 100 cu km/year since 2000, almost entirely because of increased ice discharge along accelerating glaciers in West Antarctica and the Antarctic Peninsula (Figures 10–12).

Dynamic response of outlet glaciers to ice-shelf breakup

Recent rapid changes in marginal regions of both ice sheets include regions of glacier thickening and slowdown



Ice Sheets and Ice Volume, Figure 10 Antarctica, showing rates of surface-elevation change derived from satellite radar-altimeter measurements. The graph on the *right* shows rates at which the ice-sheet mass was estimated to be changing based on radar-altimeter data (*black*), mass-budget calculations (*red*), and satellite gravity measurements (*blue*). Rectangles depict the time periods of observations (*horizontal*) and the upper and lower estimates of mass balance (*vertical*). Note that an ice loss of about 360 Gt/year ($=400 \text{ km}^3/\text{year}$) raises sea level by 1 mm/year (Figure compiled by R. Thomas and S. Manizade).

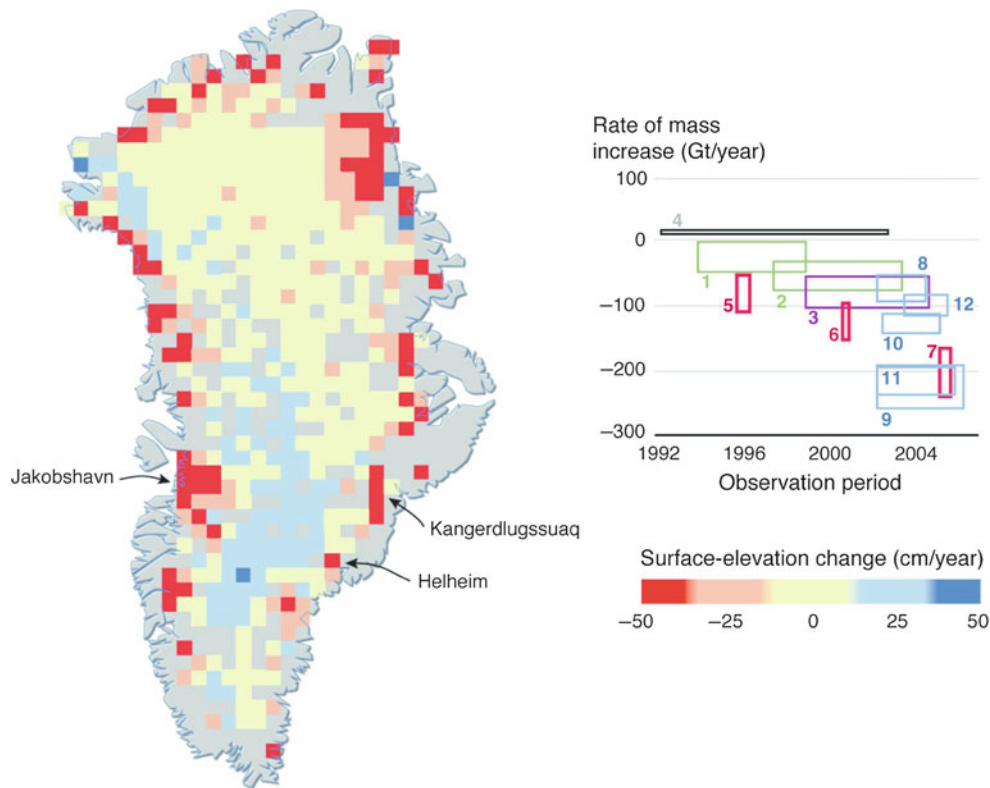
but mainly acceleration and thinning, with speeds of some glaciers more than doubling. Most of these accelerations closely followed reduction or loss of ice shelves. Such behavior was predicted 40 years ago by John Mercer, based on evidence left by previous ice ages, and by Terry Hughes, based on the surface profiles of existing glaciers flowing into the Ross Ice Shelf. Soon after, Bob Thomas found evidence for the buttressing effects of ice shelves in his measurements of ice-shelf spreading rates. But the suggestion was discounted by most glaciologists as recently as the Third IPCC Assessment Report on Climate Change in 2001, based largely on results from prevailing model simulations. Soon after this, however, total breakup of the floating ice tongue of Jakobshavn Glacier in Greenland, preceded by its very rapid thinning and followed by doubling of the glacier velocity to 14 km/year, clearly showed the sensitive link between floating ice shelves and the glaciers that flow into them (Figure 13).

Most ice shelves are in Antarctica, where they cover an area of about 1.5 million sq km with nearly all ice streams and outlet glaciers flowing into them. Several small ice shelves on the Antarctic Peninsula collapsed in the last three decades of the twentieth century, possibly related to local atmospheric warming of about 3°C over the second half of

the twentieth century. Most of these ice-shelf breakups were closely followed by large accelerations of tributary glaciers, confirming the predictions discussed above. Based on these observations, it appears that ice shelves become vulnerable to breakup if mean annual air temperature exceeds -5°C . It is also probable that ocean warming preconditions an ice shelf for breakup by thinning the ice shelf as basal melting increases, and the very rapid thinning of the Jakobshavn floating ice tongue coincided with inflow of very warm ocean waters into the glacier's fjord.

Outlook: the future

Model results predict increasing snowfall in a warming climate in Antarctica and Greenland, but only the latter has been verified by independent measurements. These studies also show significantly increased melt-water runoff from Greenland in a warming climate, partly compensated by the increased snowfall, and data from recent years suggest an increase in net loss from the surface mass balance. Consequently, because there is summer melting over $\sim 50\%$ of Greenland already, and a small temperature rise substantially increases the area of summer melting, the ice sheet is particularly susceptible to continued warming. Model studies indicate that



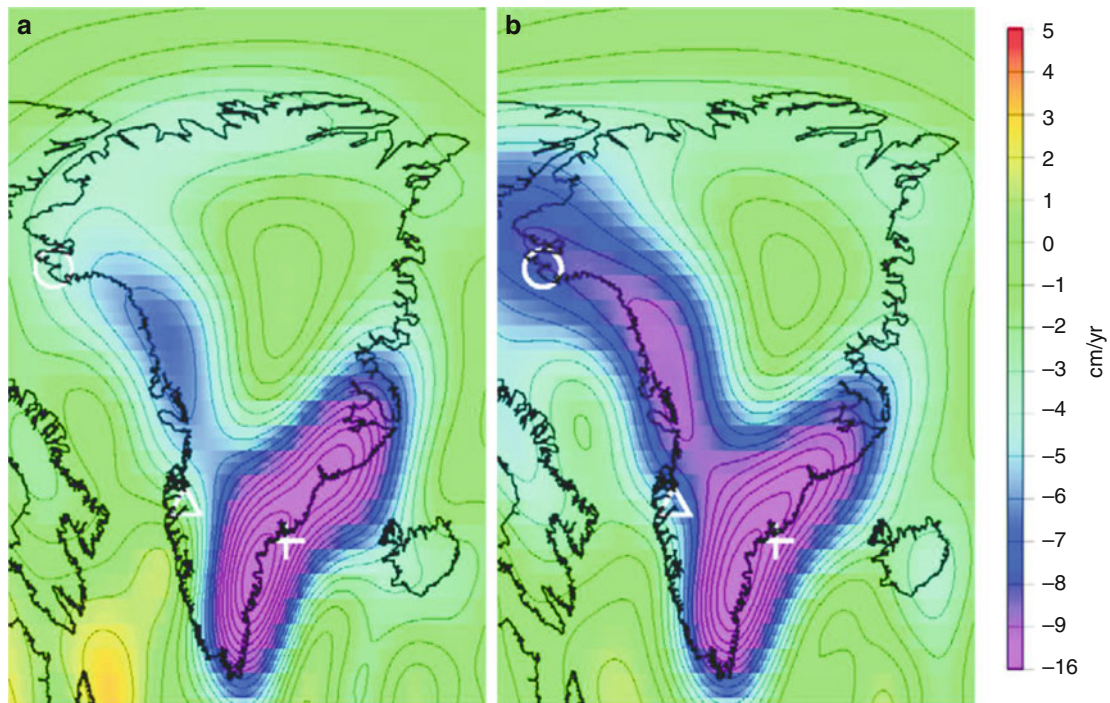
Ice Sheets and Ice Volume, Figure 11 Greenland, showing rates of surface-elevation change between the late 1990s and 2003, derived by comparing satellite and aircraft laser-altimeter surveys. The graph shows rates at which the ice-sheet mass was estimated to be changing based on satellite radar-altimeter data (black), laser-altimeter surveys (green and purple), mass-budget calculations (red), temporal changes in gravity (blue). Jakobshavn, Helheim, and Kangerdlugssuaq are fast glaciers that doubled in speed recently (Figure compiled by R. Thomas and S. Manizade).

a temperature increase over Greenland by more than 3 °C would probably result in irreversible loss of the ice sheet. Moreover, this estimate is based on imbalance between snowfall and melting, and ice loss would be accelerated by changing glacier dynamics of the type we are already observing.

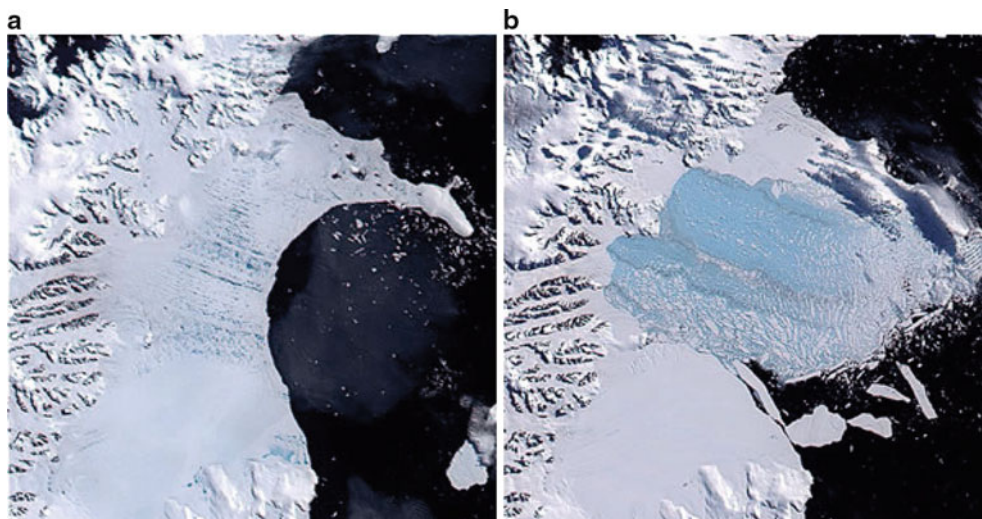
Although Greenland glacier tongues and ice shelves are vulnerable to ocean warming, there are few of them, most flow in deep troughs for only short distances inland, and central parts of the ice sheet are well protected by the fringe of near-coastal mountains. Antarctic Peninsula ice shelves are also small with comparatively little ice flowing into them, so their collapse has negligible impact on global sea level. Moreover, the largest ice shelves, in the Weddell and Ross Sea Embayments, are far to the south where it is considerably colder than the -5°C viability criterion cited above. Much of the ice draining from the West Antarctic ice sheet flows into these two ice shelves, with the remainder flowing into the Amundsen Sea to the north. Here, by contrast, the ice shelves are quite small and many have been thinning over the past decade or more, with associated acceleration of tributary glaciers that flow along very deep troughs. In particular, the speed

of *Pine Island Glacier* has roughly doubled since the 1970s, as its ice shelf thinned by as much as 5 m/year. This is a very large glacier, currently discharging more than 100 km^3 ice/year into the ocean – approximately double its “balance” value. Nearby glaciers have also accelerated, and this region is probably the most vulnerable in Antarctica to climate warming particularly if, as appears likely, this is accompanied by ocean warming.

The regions likely to experience future rapid changes in ice volume are those where ice is grounded well below sea level such as the West Antarctic ice sheet and large glaciers in Greenland like Jakobshavn Isbrae that flow into the sea through a deep channel reaching far inland. The interaction of warm ocean waters with the calving fronts and floating extensions of these glaciers represents a strong potential cause of abrupt change in the big ice sheets, and future changes in ocean circulation and ocean temperatures will very probably produce changes in ice-shelf basal melting, but the magnitude of these changes cannot currently be modeled or predicted. Moreover, calving, which can originate in fractures far back from the ice front, and ice-shelf breakup are very poorly understood. Inclusion of these processes in models will certainly yield sea-level projections



Ice Sheets and Ice Volume, Figure 12 The rate of mass loss, in cm/year water equivalent thickness, determined from monthly GRACE gravity field solutions. (a) The rate averaged between February 2003 and February 2007. (b) The rate averaged between February 2003 and June 2009 (Khan et al., 2010).



Ice Sheets and Ice Volume, Figure 13 Breakup of the Larsen B ice shelf. These are images from NASA's MODIS satellite sensor. Part of the Antarctic Peninsula is on the left. The image on the left (a) shows the shelf in late summer, with dark bluish melt ponds on the surface. The image on the right (b), collected only 5 weeks later, shows a large part of the ice shelf has collapsed, with thousands of sliver icebergs at the margins and a large blue area of ice fragments (Images: National Snow and Ice Data Center, Boulder, Colorado).

for the end of the twenty-first century that substantially exceed the projections presented in 2007 as part of the Fourth IPCC Assessment of Climate Change Report (0.28–0.42 m/year rise in sea level).

Bibliography

- Alley, R., et al., 1993. Abrupt increase in snow accumulation at the end of the Younger Dryas event. *Nature*, **362**, 527–529.
- Allison, I., et al., 2009. Ice sheet mass balance and sea level. *Antarctic Science*, **21**(5), 413–426.
- Bentley, C., et al., 2007. Ice sheets. In *Global Outlook for Ice and Snow*. Nairobi: United Nations Environment Programme, pp. 99–113.
- Khan, S. A., et al., 2010. Spread of ice mass loss into Northwest Greenland observed by GRACE and GPS. *Geophysical Research Letters*, **37**, L06501, doi:10.1029/2010GL042460, 2010.
- Pritchard, H. D., et al., 2009. Extensive dynamic thinning on the margins of the Greenland and Antarctic ice sheets. *Nature*, doi:10.1038/nature08471.
- Riva, E. M., et al., 2009. Glacial isostatic adjustment over Antarctica from combined ICESat and GRACE satellite data. *Earth and Planetary Science Letters*, **288**, 516–523.

ICEBERGS

Donald L. Murphy
International Ice Patrol, US Coast Guard,
New London, CT, USA

Synonyms

Glacier berg; Glacier ice; Ice island

Definition

Iceberg. A floating ice mass extending more than 5 m above the sea surface, which has calved from a glacier or ice shelf (World Meteorological Organization, 2007).

Ice shelf. A thick sheet of floating ice attached to land or a glacier (World Meteorological Organization, 2007).

Introduction

Long before *RMS Titanic* struck an iceberg and sank in the North Atlantic in April 1912, mariners were well aware of the beauty and danger icebergs present. The first documented iceberg report might have been provided as long ago as the sixth century by the legendary Irish monk, St. Brendan, who reported encountering floating crystal palaces during his North Atlantic voyage. It was *Titanic*, however, that brought a sense of urgency to tracking icebergs and distributing that information to mariners, a task that was assigned to the US Coast Guard International Ice Patrol (IIP). Despite impressive technological advances in shipboard radars, electronic navigational systems, and communications, ships still collide with ice. In November 2007, the cruise ship *MS Explorer* sank 20 h after striking ice off King George Island, near

Antarctica. Fortunately, there was no loss of life. While it is not entirely clear whether *Explorer* struck sea ice or an iceberg, it is clear that the danger ice poses to ships is unabated. This threat extends to other offshore activities, such as oil production and exploration. For example, during the 2008 iceberg season in the western North Atlantic, a large number of icebergs threatened offshore oil platforms near Newfoundland, Canada. Production stopped for several days, and an exploratory drill rig suspended operations and departed the iceberg danger area.

Aside from the clear menace of icebergs to maritime safety, other reasons to examine iceberg populations include their role in mass balance and freshwater flux. Ice discharge is an integral part of ice-sheet mass balance. For instance, giant (>18.5 km long) icebergs calved from the Antarctic ice shelves and glacier tongues accounted for approximately half of the mass lost by the Antarctic ice sheet during the last 25 years (Silva et al., 2006). In the Arctic, the 2005 calving of a 66.4 km² ice island from the Ayles Ice Shelf, on Ellesmere Island, Canada (Copeland et al., 2007), reduced the area of the remaining Ellesmere ice shelves by about 7.5 %. Further, the subsequent melting of the calved icebergs contributes to the flux of freshwater to the ocean. The meltwater from the giant icebergs calved from Antarctica was shown to be significant in the Southern Ocean's freshwater balance and exceeded precipitation minus evaporation (P – E) in the Weddell Sea (Silva et al., 2006). Large icebergs also have the potential to alter the dynamics of the marine ecosystem substantially (Arrigo et al., 2002).

The earliest data on icebergs were obtained from ships. These include the extensive dataset collected by the Australian and Russian vessels in the Indian and Atlantic sectors of the Southern Ocean (Jacka and Giles, 2007; Romanov et al., 2008), the pre-World War II part of IIP's western North Atlantic iceberg dataset (IIP, 1913–1946), as well as others. The vast extent and isolated nature of the ocean regions populated by icebergs make extensive and frequent shipboard observations impractical; thus, the wide array of satellite-borne, airborne, and surface-based remote sensing instruments is particularly attractive in undertaking iceberg studies.

Iceberg properties and sizes

An iceberg is a floating ice mass extending more than 5 m above the sea surface (World Meteorological Organization, 2007). It may originate from a glacier flowing directly to the sea, such as the tidewater glaciers of Greenland (e.g., Joughin et al., 2004), or from an ice shelf. An ice shelf is a thick sheet of floating ice attached to land or a glacier (World Meteorological Organization, 2007). It may be nourished by snow falling directly on its surface. The large icebergs that calve from Antarctic ice shelves are usually referred to as tabular icebergs, while the term “ice island” is more commonly used to describe icebergs from Arctic ice shelves.

Icebergs, Table 1 Iceberg size categories used in the western North Atlantic and the Antarctic

North Atlantic (Environment Canada, 2005)			Australian Antarctic Program (Jacka and Giles, 2007)	
Description	Height (m)	Length (m)	Category	Length (m)
Growler	<1	<5	1	25–100
Bergy bit	1 to <5	5 to <15	2	100–200
Small iceberg	5–15	15–60	3	200–400
Medium iceberg	16–45	61–120	4	400–800
Large iceberg	46–76	121–200	5	800–1,600
Very large iceberg	>75	>200	6	1,600–3,200
			7	>3,200

Depending on their source, icebergs may have slightly differing compositions. The icebergs calved from Greenland's glaciers tend to be mostly glacial ice, with a density approximately that of pure ice (917 kg/m^3). Thus, approximately one-seventh of the mass of the iceberg will be above the ocean's surface (Lewis et al., 1994). The large tabular icebergs calved from Antarctic ice shelves are covered in snow, which reduces their mean density (Weeks and Mellor, 1978).

Icebergs come in a remarkable array of shapes with sizes spanning the range from a piano to a Caribbean Island. The sail height and keel depth of icebergs can vary widely depending on the shape of the iceberg. Typical freeboard heights for the large tabular icebergs calved from the major Antarctic ice shelves are 30–40 m (Jacka and Giles, 2007), while icebergs in the North Atlantic may reach heights of 100 m or more.

Iceberg length is a dominant factor in selecting the appropriate remote sensing system. The Canadian Ice Service (CIS) and IIP use a size classification scheme (Table 1) that is tailored for the icebergs entering the busy shipping lanes in the western North Atlantic Ocean. It reflects the fact that small pieces of ice can severely damage a vessel. An iceberg exceeding just a few kilometers in length is very rare. This classification scheme is wholly inadequate for the massive icebergs found near Antarctica. Although there is no universally accepted size classification scheme for Antarctic icebergs, the one used by the Australian Antarctic Program (Table 1) accounts for a wider range of large icebergs. To further emphasize the size disparity, an Antarctic iceberg must measure at least 18.5 km along the long axis to be entered into the United States (US) National Ice Center's (NIC) iceberg database (NIC, 2008).

Using remotely sensed information to study icebergs

Both passive and active sensors have been used successfully in iceberg studies. Passive sensors, which rely on

natural illumination and emission, include visual and infrared systems and passive microwave sensors. Active sensors, which provide their own illumination, include radar systems such as radar altimeters and imaging radars. Each has advantages and disadvantages for areal coverage, spatial and temporal resolution, and the ability to provide data irrespective of weather and light conditions. The following discussion describes a few of the sensors that have provided remotely sensed data to advance the knowledge of the distribution and behavior of icebergs. It focuses on satellite-borne instruments because of their ability to cover wide areas, a significant advantage when monitoring icebergs. The discussion includes some of the inevitable trade-offs between resolution, an indicator of the iceberg length that can be detected, and areal coverage (swath) of the instrument. For some of the instruments, the polarization of the signal is also an important characteristic. This is particularly significant to the ability of synthetic aperture radars (SAR) to distinguish between icebergs and vessels.

Visual and infrared sensors

Satellite-borne visual and infrared sensors, such as the Advanced Very High Resolution Radiometer (AVHRR) on the National Oceanic and Atmospheric Administration's (NOAA) polar-orbiting satellites and the Operational Linescan System (OLS) on the Defense Meteorological Satellite Program (DMSP) satellites, are major sources of the NIC database of giant (>18.5 km in length) Antarctic icebergs (Long et al., 2002). This database has been used in numerous studies of Antarctic iceberg populations including an examination of the contribution of giant icebergs to the freshwater flux to the Southern Ocean (Silva et al., 2006). The nominal spatial resolution of the AVHRR is 1.1 km (NOAA, 1998), while the OLS fine-mode resolution is 0.55 km (NOAA, 2008). In both cases, the resolution is well suited to monitoring large icebergs in the Antarctic. The data are of relatively low cost and are available in near real time. Unfortunately, clouds and low-light conditions during the polar winter often hamper visual and infrared sensors, which may result in data gaps.

Passive microwave systems

Because they operate in the microwave range, passive microwave systems penetrate most clouds giving them a major advantage over the visible and infrared sensors. The 85 GHz channel on the special sensor microwave imager (SSM/I) on the DMSP satellite has been used to track large Antarctic icebergs (Hawkins et al., 1991). With a 16 km by 14 km footprint, this SSM/I channel is capable of detecting icebergs of several tens of km in length and larger.

The Advanced Microwave Scanning Radiometer (AMSR-E) on the Aqua satellite, launched in 2002 and functioned until 2011, was a significant improvement over SSM/I, particularly because it has much-improved spatial

resolution. The AMSR-E's 89 GHz channel had a much smaller footprint (6 km by 4 km), allowing it to detect smaller icebergs. It has been used to track 10 km icebergs in the Antarctic (Blonski and Peterson, 2006). As with the visual and infrared sensors, passive microwave sensors are adequate for tracking the large icebergs of the Antarctic but have limited use for iceberg tracking in the northern hemisphere.

Active microwave systems

Active microwave systems are becoming the dominant remote sensing instruments in the study of icebergs in both hemispheres. They include radar scatterometers and imaging radars. Many of the instruments offer finer resolution than visual, infrared, and passive microwave systems. More importantly, these systems provide their own illumination and are not weather dependent.

Radar scatterometers were designed to measure winds over the ocean from space, but because of the high radar backscatter of icebergs, they are useful for iceberg tracking as well. The spatial resolution varies according to the specific instrument but ranges from 2.225 to 8.9 km/pixel (Long et al., 2002). The data from five scatterometers were used to study the Antarctic iceberg population continuously from 1992 by Ballantyne and Long (2002). They suggest that the increase in the number of Antarctic icebergs observed in recent years is due to the improved detection and tracking technology. Stuart et al. (2008) maintain an online database of the iceberg locations derived from the scatterometer data.

The 1978 launch of the US National Aeronautics and Space Administration's (NASA) SEASAT ushered in the era of ocean-observing synthetic aperture radars (SAR). Although SEASAT provided data for only a few months, it paved the way for a progression of SAR satellites used to monitor icebergs over a wide range of sizes in both hemispheres. Since the early 1990s, imaging radars launched by the European Space Agency (ESA), the Canadian Space Agency (CSA), and others have become increasingly capable and flexible. These instruments offer numerous beam modes with various swaths and resolutions, some with the ability to detect 30 m icebergs (Power et al., 2001). The all-weather capability and wide range of iceberg sizes that can be detected make satellite-borne SARs valuable for both operations and research.

Because small icebergs pose the greatest threat to ships, the ability to detect them is particularly significant to maritime-safety operations. Such icebergs are large enough to inflict serious damage, yet small enough to elude detection in high sea states. Two challenges dominate the operational use of SARs to detect icebergs: first, the trade-off between resolution and coverage (swath) and, second, the ability to distinguish between icebergs and vessels. CSA's RADARSAT-1 is a single-polarization radar. For high incidence angles ($>35^\circ$), it detects icebergs on the order of the resolution of the particular SAR mode. For example, in one mode, which has a 150 km swath,

icebergs 30 m or larger can be detected (Power et al., 2001). In another mode, the swath increases to 300 km, but the resolution becomes 50 m. Single-polarization images do not always provide reliable ability to distinguish between ships and icebergs. The alternating polarization mode on the Advanced SAR (ASAR) on ESA's ENVISAT, launched in 2002, permits analysis of different polarization responses of ships and icebergs (Howell et al., 2004). The increased flexibility of choosing the signal polarization offered by the RADARSAT-2 improves the discrimination skills (Scheuchl et al., 2004).

Satellite-borne SAR data are providing a rich source of information for the study of iceberg populations throughout the world's oceans. In addition, because they are capable of detecting a wide range of sizes, they give a better representation of the populations. For example, the SARs on ESA's European Remote Sensing Satellites (ERS-1, ERS-2) and RADARSAT-1 were used to study the iceberg distribution in the Eurasian Arctic. Icebergs present there are much smaller than the Antarctic, with an average size of about 100 m. SAR images showed significant iceberg production from the Renown Glacier in Franz Josef Land (Alexandrov et al., 2003). SAR data in the Antarctic have provided another resource for the extensive iceberg database being maintained at the NIC (National Ice Center, 2008).

Two trends are emerging from the identification and tracking of icebergs using SAR data. First, the data sets are becoming so large that automated systems are required to conduct consistent and efficient analysis. Computer-based techniques have been developed to locate and identify Antarctic icebergs using SAR images (Silva and Bigg, 2005), and the system also retains the iceberg shape to facilitate tracking using sequential images. Second, the process of interpreting SAR images sometimes leads to ambiguities. For example, distinguishing between small icebergs and small multiyear sea ice floes can be difficult even for a skilled analyst. Small icebergs and small vessels are also difficult to distinguish.

Summary

The last few decades have witnessed remarkable improvements in the ability of satellite-borne sensors to detect, track, and measure icebergs of wide size range. The earliest capability was limited to detecting large Antarctic icebergs in favorable weather and light conditions. Now, numerous sensors, many unencumbered by weather and light conditions, are used in both hemispheres to track icebergs as small as a few tens of meters.

An enormous increase in the amount of available data, particularly from SARs, has led to the development of automated image-analysis techniques to detect icebergs. While many of these systems have proven successful, it is widely agreed that the human analyst cannot yet be removed entirely from the process. In addition, there is still much work to be done comparing remotely sensed data with surface observations.

Finally, it is likely that human activity in ice-populated waters will increase due to resource exploitation in the ice-diminishing Arctic and growing public interest in ecotourism. This will create new challenges for many of the world's operational ice centers. Today, 20 countries (World Meteorological Organization, 2006) have organizations that provide ice services. Their success at meeting future demands hinges on the effective and efficient use of remotely sensed data.

Bibliography

- Alexandrov, V. Y., Sandven, S., and Kloster, K., 2003. Iceberg identification in the Eurasian Arctic using SAR images. In *Proceedings 2003 IEEE International Geoscience and Remote Sensing Symposium*, Vol. 4, pp. 2798–2801.
- Arrigo, K. R., van Dijken, G. L., Ainley, D. G., Fahnestock, M. A., and Markus, T., 2002. Ecological impact of a large Antarctic iceberg. *Geophysical Research Letters*, **29**(7), 1104.
- Ballantyne, J., and Long, D. G., 2002. A multidecadal study of the number of Antarctic icebergs using scatterometer data. In *Proceedings 2002 IEEE International Geoscience and Remote Sensing Symposium*, Vol. 5, pp. 3029–3031.
- Blonski, S., and Peterson, C. A., 2006. Antarctic iceberg tracking based on time series of aqua AMSR-E microwave temperature measurements. *EOS, Transactions American Geophysical Union*, **87**(52), Fall Meeting Supplement, Abstract # IN33B–1352.
- Copeland, L., Mueller, D. R., and Weir, L., 2007. Rapid loss of the Ayles ice shelf, Ellesmere island. *Canada Geophysical Research Letters*, **34**, L21501.
- Environment Canada, 2005. *Manual of Standard Procedures for Observing and Reporting Ice Conditions (MANICE)*, 9th edn. Ottawa: Canadian Ice Service, Meteorological Service of Canada, Environment Canada.
- Hawkins, J. D., Laxon, S., and Phillips, H., 1991. Antarctic tabular iceberg multi-sensor mapping. In *Proceedings 1991 International Geoscience and Remote Sensing Symposium, Remote Sensing: Global Monitoring for Earth Management*, Vol. 3, pp. 1605–1608.
- Howell, C., Youden, J., Lane, K., Power, D., Randell, C., and Flett, D., 2004. Iceberg and ship discrimination with ENVISAT multi-polarization ASAR. In *Proceedings 2004 IEEE International Geoscience and Remote Sensing Symposium*, Vol. 1, pp. 113–116.
- International Ice Patrol Annual Reports, 1913–1946. *International Ice Observation and Ice Patrol Service in the North Atlantic Ocean*. Bulletins 1–32. Groton, CT: U.S. Coast Guard International Ice Patrol.
- Jacka, T. H., and Giles, A. B., 2007. Antarctic iceberg distribution and dissolution from ship-based observations. *Journal of Glaciology*, **53**(182), 341–356.
- Joughin, I., Abdalati, W., and Fahnestock, M., 2004. Large fluctuations in speed on Greenland's Jakobshavn Isbrae glacier. *Nature*, **432**, 608–610.
- Lewis, E. O., Livingstone, C. E., Garrity, C., and Rossiter, J. R., 1994. Properties of snow and ice. In Haykin, S., Lewis, E. O., Raney, R. K., and Rossiter, J. R. (eds.), *Remote Sensing of Sea Ice and Icebergs*. New York: Wiley.
- Long, D. C., Ballantyne, J., and Bertoia, C., 2002. Is the number of Antarctic icebergs really increasing? *EOS Transactions American Geophysical Union*, **83**(42), 469.
- National Ice Center, 2008. *Antarctic Icebergs*. Retrieved 13 June 2008 from <http://www.natice.noaa.gov/products/iceberg/index.htm>.
- National Oceanographic and Atmospheric Administration, 1998. *NOAA Polar Orbiter data User's Guide*. Retrieved 14 June 2008 from <http://www2.ncdc.noaa.gov/docs/podug/index.htm>.
- National Oceanographic and Atmospheric Administration, 2008. *DMSP Sensors*. Retrieved 14 June 2008, from <http://www.ngdc.noaa.gov/dmsp/sensors.html>.
- Power, D., Youden, J., Lane, K., Randell, C., and Flett, D., 2001. Iceberg detection capabilities of RADARSAT synthetic aperture radar. *Canadian Journal of Remote Sensing*, **27**(5), 476–486.
- Romanov, Y. A., Romanova, N. A., and Romanov, P., 2008. Distribution of icebergs in the Atlantic and Indian Ocean sectors of the Antarctic region and its possible links with ENSO. *Geophysical Research Letters*, **35**, L02506.
- Scheuchl, B., Flett, D., Caves, R., and Cumming, I., 2004. Potential of RADARSAT-2 data for operational sea ice monitoring. *Canadian Journal of Remote Sensing*, **30**(3), 448–461.
- Silva, T. A. M., and Bigg, G. R., 2005. Computer-based identification and tracking of Antarctic icebergs in SAR images. *Remote Sensing of Environment*, **94**(3), 287–297.
- Silva, T. A. M., Bigg, G. R., and Nicholls, K. W., 2006. Contribution of giant icebergs to the Southern Ocean freshwater flux. *Journal of Geophysical Research*, **111**, C03004.
- Stuart, K., Lambert, B., Ballantyne, J., and Long, D. G., 2008. *The Antarctic Iceberg Tracking Database, 1978 & 1992–2008*, Brigham Young University. Provo: UT. Retrieved 13 June 2008 from <http://www.scp.byu.edu/data/iceberg/database1.html>.
- Weeks, W. F., and Mellor, M., 1978. Some elements of iceberg technology. In Hussein, A. A. (ed.), *Iceberg Utilization, Proceedings First International Conference*. New York: Pergamon, pp. 45–98.
- World Meteorological Organization, 2006. *Sea-Ice Information Services in the World*. WMO-No. 574, 3rd edn. Geneva: World Meteorological Organization. <http://www.jcomm-services.org/documents.htm?parent=136>.
- World Meteorological Organization, 2007. *Sea-ice nomenclature. WMO Sea Ice Nomenclature*, draft, ver. 1.1. Geneva: World Meteorological Organization. <http://www.jcomm-services.org/documents.htm?parent=136>.

Cross-references

Ice Sheets and Ice Volume
Microwave Radiometers
Observational Systems, Satellite
Radar, Synthetic Aperture

INTERNATIONAL COLLABORATION

Lisa Robock Shaffer

MC 0553 Rady School of Management, University of California, San Diego, La Jolla, CA, USA

Synonyms

Bilateral or multilateral collaboration; International cooperation

Definition

A variety of forms of participation in remote sensing missions by organizations or individuals from more than one country. This includes space-based missions whose hardware includes components from different countries, as well as science team participation by persons

from countries other than the country of the Principal Investigator. International collaboration can also take the form of multilateral organizations in which program planning, education and outreach, calibration and validation, and other aspects of program coordination take place.

Introduction

International collaboration in remote sensing can take place in any or all elements of a typical mission: mission planning, spacecraft bus, instrument payload, launch vehicle, operations, data acquisition, and data analysis. From the simple sharing of an image acquired from space to the design of an integrated mission involving multiple partners, the history of space-based Earth observations provides examples of all levels of international engagement.

Data Sharing Across Borders: The simplest form involves someone in one country using data from a mission provided by another country. It can involve the transmission of data received centrally and distributed through terrestrial means (postal mail or electronic transmission). In the early history of Earth observations, this was typical – a magnetic tape or a copy of a photograph or other image would be created and shipped to an investigator in another country, after a competitive process of selecting a scientific research team through an international call for proposals. For some missions where there is interest in the data for purposes beyond the scientific objectives of the sponsoring agency, data have been provided on varying terms from no charge to cost of reproduction to commercial prices, sometimes without restriction and sometimes with very specific constraints on how the data could be used and by whom. Data from US weather satellites is the least restricted, and missions that have commercial dimensions such as SeaWiFS and SPOT have the most complex [data access](#) policies.

Data sharing can involve direct reception of data from one country's satellite received at a ground station in another country. The earliest form of such collaboration is probably the US weather satellites in the early 1960s. President Kennedy decreed that all the data would be openly transmitted on a global basis, and the US Government provided technical information and assistance to enable meteorological agencies around the world to receive data from the Automated Picture Transmission (APT) system. Data from NOAA satellites continues to be openly transmitted without a fee to this day, with technical support available through the World Meteorological Organization, to enable all nations, regardless of their economic status, to have vital weather information. A more complex version of this kind of cooperation continues today with remote sensing satellites including the US Landsat and French SPOT satellite series, in which ground station operators acquire data directly from the satellite but on a scheduled, fee-based arrangement.

Instrument Provision: Some satellite payloads include instruments provided by foreign space agencies. These

can be selected through a competitive process where instrument providers submit proposals to the mission lead agency, such as was done for the original NASA Earth Observing Satellite platforms and NSCAT flying on ADEOS, or specific instrument contributions can be solicited from specific providers and negotiated on a bilateral basis, such as the Search and Rescue COSPAS/SARSAT packages that fly on the polar-orbiting meteorological satellites.

Joint Missions: When there is shared planning from the inception of a mission, we refer to the type of collaboration as a joint mission. The US-French Topex/Poseidon and JASON satellites are an example, as are missions undertaken by the European Space Agency and EUMETSAT, in which the member states develop the plans for both the jointly funded elements of the payload and any national contributions as part of the mission design from the start. The Tropical Rainfall Monitoring Mission, jointly developed by NASA and NASDA is another example.

Launch and Operations: Sometimes the international collaboration comes in the form of operations support, such as provision of a launch and support for operations. The Pegasus launch of an SAC mission for Argentina and the Ukraine launch vehicle provided for GRACE are examples. In both cases, the missions themselves are also international, SAC being a collaboration between NASA and CONAE (Argentina) and GRACE involving NASA and the German space agency, DLR.

Planning and Coordination: There are several international organizations that have played a critical role in stimulating international collaboration in remote sensing. These include the World Meteorological Organization (WMO), whose coordination of meteorology predates the existence of space-based observing systems and which has been essential in the expansion of international cooperation into the satellite era. The Coordination of Geostationary Meteorological Satellite (CGMS) group (which became the Coordination Group for Meteorological Satellites) was instrumental in developing payload harmonization and backup operational arrangements for the major geostationary weather satellite systems. The Committee on Earth Observation Satellites (CEOS) was established in 1984 and has provided a venue for technical working groups on such topics as sensor calibration and validation, data cataloging, and format standards, as well as regularly convening senior space agency officials to discuss long-range mission plans and coordinate educational outreach and develop advocacy strategies to secure and expand public support and funding for space-based Earth observations.

As a result of the work of these groups and others, in 2003 the first Earth Observation Summit was convened at ministerial level and established a process that has resulted in creation of the Global Earth Observing System of Systems (GEOSS) and the Group on Earth Observations (GEO), formed in 2005 to carry out the 10 year GEOSS implementation plan.

Summary

As the need to observe and understand the planet increases in importance, and as the recognition of our global interdependence grows, it is only natural that the advanced technologies used to observe the Earth from space would stimulate international collaboration in many forms. The technological and scientific challenges require the best and brightest from across the planet to address the critical challenges of climate change and sustainability.

Bibliography

Committee on Earth Observation Satellites (CEOS), <http://www.ceos.org/>
 EUMETSAT, <http://www.eumetsat.int/Home/index.htm>
 European Space Agency, Earth Observations, [http://www.esa.int/Group_on_Earth_Observations_\(GEO\)](http://www.esa.int/Group_on_Earth_Observations_(GEO)), <http://earthobservations.org/>
 NASA Office of External Relations, Science Division, <http://www.hq.nasa.gov/office/oer/science.html>
 NOAA/NESDIS International and Interagency Affairs Office, <http://www.noaa.gov/>

Cross-references

Data Policies
 Global Climate Observing System
 Global Earth Observation System of Systems (GEOSS)
 Global Land Observing System
 Policies and Economics
 Public-Private Partnerships
 Remote Sensing, Historical Perspective

IONOSPHERIC EFFECTS ON THE PROPAGATION OF ELECTROMAGNETIC WAVES

Attila Komjathy
 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

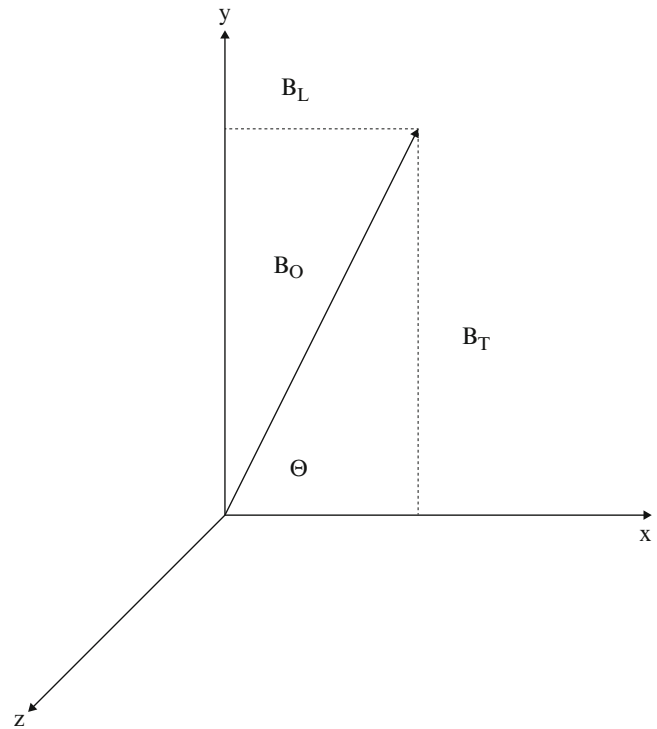
Ionospheric wave specification or wave propagation;
 Propagation of electromagnetic waves in the ionosphere

Definition

- *Appleton-Hartree Formula*. the equation describing the medium that is electrically neutral with constant magnetic field impressed upon it.
- *Ground and Space-Based Remote Sensing Techniques*. instruments placed on the ground or onboard satellites measuring critical ionospheric parameters and integrated electron densities (i.e., total electron content or TEC).

Introduction to ionospheric propagation

The effect of the ionosphere on electromagnetic waves propagation (see “*Electromagnetic Theory and Wave Propagation*”) can only partially be described by simple dispersion. To adequately describe the complete behavior



Ionospheric Effects on the Propagation of Electromagnetic Waves, Figure 1 System of orthogonal axes x, y, z (After Davies, 1990).

of radio waves in the ionosphere, it is important to realize that the “ionosphere is a partially ionized, spherically stratified plasma with a wide spectrum of non-uniformly spaced irregularities, upon which is imposed a non-uniform magnetic field” (Hunsucker, 1991).

The complex refractive index of the ionosphere as a magnetoionic medium was derived by the number of scientists, but the name most commonly associated with the theory is Sir Edward Appleton who was first to point out that a plane-polarized wave would be split into two opposite rotating circularly polarized waves by the magnetized plasma (Hunsucker, 1991). In 1931, Hartree suggested the inclusion of the Lorentz polarization term after which event the complex refractive index was often referred to as the Appleton-Hartree formula.

The detailed derivation of the formula can be found in Davies (1966, 1990) and Komjathy (1997). First, we have to apply Maxwell’s equations to the wave, and secondly we have to impose the properties of the medium, the so-called constitutive relations. *The Appleton-Hartree magnetoionic theory* applies to a medium that is electrically neutral with no resultant space charge and an equal number of electrons and positive ions upon which a constant magnetic field (see “*Earth Magnetic Field*”) is impressed, and the effect of positive ions on the wave is

negligible. Let us consider a plane electromagnetic wave travelling in the x direction of the orthogonal coordinate system displayed in Figure 1.

Let us also consider a uniform external magnetic field that lies in the x-y plane and makes an angle θ with the -direction of propagation. The complex refractive index n is given by the Appleton-Hartree magnetoionic dispersion equation (see e.g., Langley, 1996; Davies, 1966, 1990; Hunsucker, 1991; Hall and Barclay, 1989):

$$n^2 = 1 - \frac{X}{(1 - jZ) - \left[\frac{Y_T^2}{2(1 - X - jZ)} \right] \pm \left[\frac{Y_T^4}{4(1 - X - jZ)^2} + Y_L^2 \right]^{1/2}}, \quad (1)$$

where n is the complex refractive index ($\mu - j\chi$) with μ being the real part and χ being the imaginary part. Furthermore:

$$X = \frac{\omega_N^2}{\omega^2} = \frac{f_N^2}{f^2}, \quad (2)$$

$$Y = \frac{\omega_H}{\omega} = \frac{f_H}{f}, \quad (3)$$

$$Y_L = \frac{\omega_L}{\omega}, \quad Y_T = \frac{\omega_T}{\omega}, \quad (4)$$

$$Z = \frac{\omega_c}{\omega}, \quad (5)$$

where ω (radian/s) is the angular frequency of the “exploring wave” f (Hz), and ω_c (radian/s) is the angular collision frequency between electrons and heavier particles; ω_N is the angular plasma frequency with $\omega_N^2 = \frac{Ne^2}{\epsilon_0 m}$ with electron density N ($1/\text{m}^3$), electron charge e ($1.6 \cdot 10^{-19}$ C), permittivity of free space ϵ_0 ($8.8542 \cdot 10^{-12}$ F/m), and electron mass m ($9.1095 \cdot 10^{-31}$ kg); ω_H is the angular gyrofrequency and $\omega_H = \frac{B_o |e|}{m}$ (radian/s) with the electromagnetic field strength B_o (Wb/m²); ω_L is the longitudinal angular gyrofrequency and $\omega_L = \frac{B_o |e|}{m} \cos \Theta$ (radian/s); ω_T is the transverse angular gyrofrequency and $\omega_T = \frac{B_o |e|}{m} \sin \Theta$ (radian/s).

When collisions are negligible (i.e., $Z \approx 0$),

$$n^2 \cong \mu^2 = 1 - \frac{2X(1 - X)}{2(1 - X) - Y_T^2 \pm \left[Y_T^4 + 4(1 - X)^2 Y_L^2 \right]^{1/2}}, \quad (6)$$

then according to the magnetoionic theory, a plane-polarized electromagnetic wave will be split into two characteristic waves: an ordinary wave which approximates the behavior of a wave propagating without an imposed magnetic field displayed with a sign “+” in Equation 6, and the wave with the sign “−” is called extraordinary wave.

The *Appleton-Hartree formula* assumes that the electron collision frequency ν is not dependent on the electron velocity. In the lower D and E regions of the ionosphere where the collision frequency ν is comparable with the wave frequency ω , this assumption no longer holds. To take this effect into account it is necessary to generalize the magnetoionic theory. Now, let us go back to Equation 1, and for the sake of simplicity, we set the magnetic field strength components $Y_T = Y_L = 0$, so that

$$n^2 \cong 1 - \frac{X}{1 - iZ} = 1 - \frac{X}{1 + Z^2} - \frac{iXZ}{1 + Z^2}. \quad (7)$$

In Equation 7, after considering the real part of the refractive index only, using the binomial expansion for refractive index n , and then integrating it along the line of sight of the radio signal, we get for the ionospheric delay

$$d_{ion} \cong 40.3 \cdot \frac{TEC}{f^2 + \nu^2}, \quad (8)$$

where ν is the collision frequency (Hz). Equation 8 is very similar to Equation 7 with the difference of the inclusion of the term ν^2 . The collision frequency ν becomes comparable with the microwave frequencies in the D layer where it can be as high as 10^9 (Hz) (Komjathy, 1997). Above the D layer, ν has a value about 10^4 (Hz). Using the largest D layer electron density at the high solar activity time ($1.3 \cdot 10^9 \text{ (m}^{-3}\text{)}$) (Bilitza, 1990) and assuming that the D layer ranges from 50 up to about 90 km, it turns out that by neglecting the collision frequency, in a worst case, we are introducing an error into the ionospheric delay using microwave frequencies at the 0.05 mm level. It is interesting to point out that the inclusion of the collision frequency ν actually reduces the refracting properties of the medium.

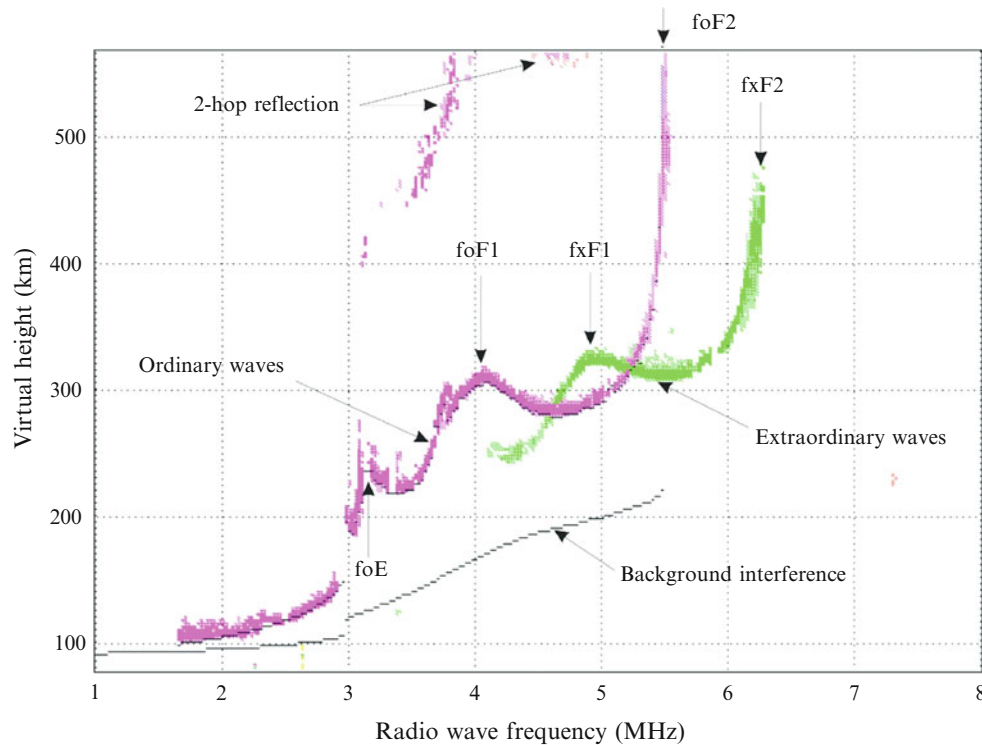
From Equation 7, we can now define the vertical incidence reflection which takes place when $n = 0$, that is,

$$1 - \left(\frac{f_N^2}{f^2} \right) = 0, \quad (9)$$

where f_N (Hz) is the plasma frequency at which a slab of neutral plasma with density N naturally oscillates after the electrons have been displaced from the ions and are allowed to move freely. f_N can also be expressed as

$$f_N^2 = \frac{Ne^2}{4\pi^2 \epsilon_0 m}, \quad (10)$$

where the individual parameters are described earlier in this section. For earth’s ionosphere with e.g., $N = 10^{12} \text{ (m}^{-3}\text{)}$, a typical value for f_N is about 8.9 MHz (Bassiri and Hajj, 1993). The condition for a wave to be reflected at vertical incidence is $f = f_N$ which is the physical principle that will make it possible to use the techniques introduced in the following sections for ionospheric studies in geodesy and remote sensing.



Ionospheric Effects on the Propagation of Electromagnetic Waves, Figure 2 Sample ionosonde data.

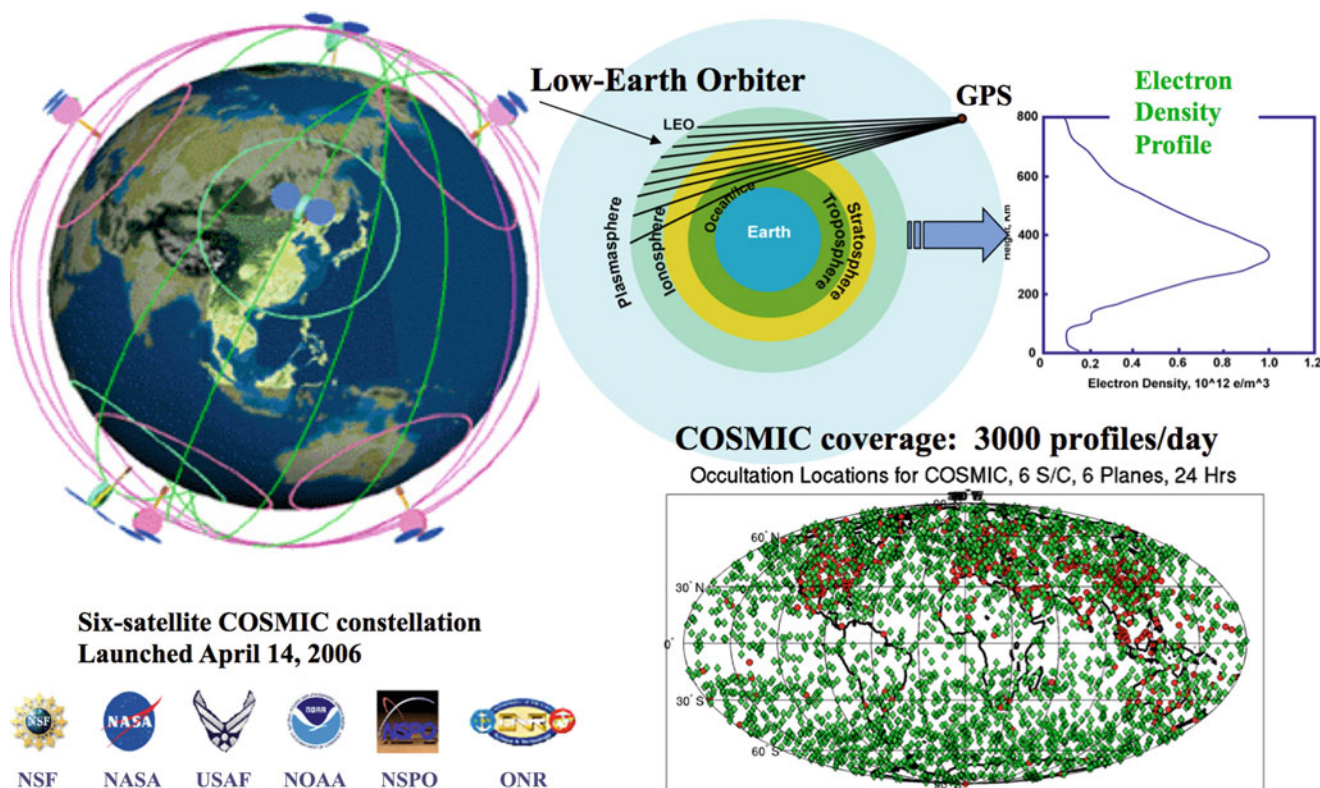
Ground-based techniques for ionospheric remote sensing

In section “Introduction to Ionospheric Propagation,” we introduced the concepts of the *refractive index* for the *magnetoionic medium*, *plasma frequency*, and the *electromagnetic wave* refracting at vertical incidence (see Equation 9). Based on these principles, we will now discuss a few examples for ground- and space-based techniques as primary ionospheric remote sensing tools to measure critical ionospheric parameters and integrated electron densities (i.e., total electron content or TEC).

Ionosonde

Ionosondes function by emitting high frequency radio waves, sweeping from lower frequency to the higher, to measure the time required for the signal to travel and return from the refracting ionospheric layer. The radio-frequency pulse travels more slowly (group velocity) in the ionosphere than in free space; therefore, the virtual height is recorded instead of the true height. For frequencies approaching the maximum plasma frequency in a particular layer, the virtual height tends to become infinity because the wave has to travel a finite distance at effectively zero speed. Ionograms can

provide the relationship between the radio wave frequency and virtual height of the reflecting ionospheric layer. From the ionograms, the characteristic values of virtual heights $h_p E$, $h_p F1$, and $h_p F2$ and critical (penetration) frequencies $f_o E$, $f_o F1$, and $f_o F2$ can be scaled manually or digitally. Modern ionosondes (digisondes) routinely scale ionograms. We present a sample ionogram in Figure 2 provided by MIT’s Haystack Observatory. The data was observed with the Millstone Hill Digisonde. The radio wave frequency is plotted against the virtual height of the reflecting layer. We can clearly see that the virtual height steadily increases with the frequency up to the critical (penetration) frequency. In Figure 2, the critical frequencies $f_o E$, $f_o F1$, and $f_o F2$ can be obtained by taking a frequency reading when the virtual height has a local minimum ($f_o E$, $f_o F1$) or when it tends to approach infinity ($f_o F2$). In the figure, we can identify two curves representing the ordinary and extraordinary waves. The curves with the higher critical frequencies $f_x F1$ and $f_x F2$ are the extraordinary waves. The ordinary and extraordinary waves can be derived by using Equation 6 and then using $n = 0$ as a criterion for reflection at an ionospheric layer. After solving the equation, we will get separate solutions for the ordinary and the extraordinary waves. We can also identify a curve at the bottom of the plot which can be



Ionospheric Effects on the Propagation of Electromagnetic Waves, Figure 3 Illustration for the COSMIC concept and constellation.

attributed to background interference. The so-called 2-hop reflection is due to the twice-reflected waves from the ionospheric layers with an intermediate ground reflection (Davies, 1990).

Incoherent backscatter radar

The incoherent scatter radar is the most powerful ground-based technique for the study of the earth's ionosphere and the interactions with the upper atmosphere, magnetosphere, and the interplanetary medium (see "Radars"). The technique is based on the radar principle which is the technique for detecting and studying remote targets by transmitting a radio wave in the direction of a target and observing the reflection of the wave. The target of any incoherent scatter radar is the electrons comprising the ionosphere. Since the amount of energy scattered by each electron is well known, the strength of the echo received from the ionosphere measures the number of electrons in the scattering volume and thus the electron density. The width of the spectrum is a measure of temperature of the ionosphere which can be different for ions and electrons. The shape of the spectrum is a sensitive function of the ratio of the electron and ion temperatures. Since the mixture of ions and electrons (also known as plasma) is constantly in motion in addition to the thermal motion, an overall shift of the

spectrum can be detected from which the speed of the ions and electrons can be inferred (Komjathy, 1997).

Space-based remote sensing

We seem to be in the midst of a revolution in ionospheric remote sensing driven by the abundance of ground and space-based GPS receivers, new UV remote sensing satellites, and the advent of data assimilation techniques for space weather. The success of GPS/MET inspired a number of other radio occultation missions (see "GPS, Occultation Systems") to provide ionospheric measurements over oceanic regions, including the Argentine Satellite de Aplicaciones Cientificas-C (SAC-C), the US-funded Ionospheric Occultation Experiment (IOX), and Germany's Challenging Minisatellite Payload (CHAMP) (Jakowski et al., 2007). The joint US/Taiwan Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC; <http://cosmicio.cosmic.ucar.edu/cdaac/index.html>), a new constellation of six satellites (see "Observational Systems, Satellite"), nominally provides up to 3,000 ionospheric occultations per day (Figure 3). The COSMIC 6-satellite constellation was launched in April 2006. COSMIC now provides an unprecedented global coverage of GPS occultation measurements (between 1,400 and 2,300 good soundings per day as of March 2009, see "Limb Sounding,

Atmospheric”), each of which yields electron density information with ~ 1 km vertical resolution. Calibrated measurements of ionospheric delay (total electron content) suitable for input into assimilation models are currently made available in near real time (NRT) from the COSMIC with a latency of 30–120 min. Similarly, NRT TEC data are available from two worldwide NRT networks of ground GPS receivers (~ 75 5 min sites and ~ 125 hourly sites, operated by JPL and others). The combined ground and space-based GPS datasets provide a new opportunity to more accurately specify the three-dimensional ionospheric density with a time lag of only 15–120 min. With the addition of the vertically resolved occultation data, the retrieved profile shapes are expected to model the hour-to-hour ionospheric “weather” much more accurately. JPL has begun integrating COSMIC-derived TEC measurements with ground-based GPS TEC data and assimilating these data into models such as JPL/USC Global Assimilative Ionospheric Model (GAIM) (originally sponsored by the US Department of Defense) so that three-dimensional global electron density structures and ionospheric drivers can be estimated.

The arrival of global, continuous datasets from diverse sources such as over 1,200 GPS receivers across the globe, satellite–satellite crosslink occultations, digisonde profiles, in situ satellite density measurements, and ultraviolet (UV) airglow measurements from low-Earth orbiters fueled the need to create models which are entirely data driven such as the persistence-driven thin shell model GIM (e.g., Mannucci et al., 1998; Komjathy et al., 2005) and various tomographic models such as MIDAS (Dear and Mitchell, 2006), EDAM (Angling and Cannon, 2004), and others (e.g., Rius et al., 1997). Further approaches, such as IDA3D (Bust and Crowley, 2007; Bust and Mitchell, 2008), attempt to incorporate data and ionospheric physics by use of models to define the a priori state of the ionosphere. In spite of high success in regions with excellent data coverage, these techniques are limited in forecasting ability. Newly emerged assimilative models involve first-principle physical models and yet also take in data, like the tomographic models, in an attempt to merge the benefits of both techniques. JPL/USC GAIM (e.g., Pi et al., 2003, 2004; Hajj et al., 2004; Wang et al., 2004; Mandrake et al., 2005), USU GAIM (e.g., Schunk, 2002; Scherliess et al., 2006), and the Fusion Numeric’s assimilation model (e.g., Khattatov et al., 2004; Angling and Khattatov, 2006) have all been developed to address this need with significant variations in the forward modeling and data assimilation approaches to better characterize ionospheric specifications.

Conclusion

The advent of real-time global ground- and space-based measurements is expected to revolutionize the accuracy of ionospheric propagation specification, nowcast, and forecast. Recently, the NOAA Space Weather Prediction Center developed a new data assimilation product

(Fuller-Rowell, 2005) that characterizes ionospheric TEC over the United States. In the next years, the real-time characterization of the global ionosphere is expected to become a standard product. This characterization will rely heavily on data from GPS measurements, but it will be enhanced by the real-time measurements from other sensors such as ionosondes or incoherent backscatter radars. As technology advances, societies of tomorrow are expected only to increase their need for highly accurate communications and navigation systems. Through collecting new data and finding new ways of analyzing ground and space-based GPS data to minimize signal propagation errors, scientists and operators will be sure to meet these future needs.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Angling, M. J., and Cannon, P. S., 2004. Assimilation of radio occultation measurements into background ionospheric models. *Radio Science*, **39**, RS1S08, doi:10.1029/2002RS002819.
- Angling, M. J., and Khattatov, B., 2006. Comparative study of two assimilative models of the ionosphere. *Radio Science*, **41**, RS5S20, doi:10.1029/2005RS003372.
- Bassiri, S., and Hajj, G. A., 1993. Higher-order ionospheric effects on the global positioning system observables and means of modeling them. *Manuscripta Geodetica*, **18**, 280–289.
- Bilitza, D. (ed.), 1990. *International Reference Ionosphere 1990*. Lanham, MD: National Space Science Data Center/World Data Center A for Rockets and Satellites. Report number NSSDC/WDC-A-R&S 90–22.
- Bust, G. S., and Crowley, G., 2007. Tracking of polar cap ionospheric patches using data assimilation. *Journal of Geophysical Research*, **112**, A05307, doi:10.1029/2005JA011597.
- Bust, G. S., and Mitchell, C. N., 2008. History, current state, and future directions of ionospheric imaging. *Reviews of Geophysics*, **46**, RG1003, doi:10.1029/2006RG000212.
- Davies, K., 1966. *Ionospheric Radio Propagation*. New York: Dover.
- Davies, K., 1990. *Ionospheric Radio*. London: Peregrinus.
- Dear, R. M., and Mitchell, C. N., 2006. Ionospheric imaging at mid-latitudes using both GPS and ionosondes. *Journal of Atmospheric and Solar-Terrestrial Physics*, **69**(2007), 817–825.
- Fuller-Rowell, T., 2005. USTEC: a new product from the Space Environment Center characterizing the ionospheric total electron content. *GPS Solutions*, **9**(3), 236–239.
- Hajj, G. A., Wilson, B. D., Wang, C., Pi, X., and Rosen, I. G., 2004. Data assimilation of ground GPS total electron content into a physics-based ionospheric model by use of the Kalman filter. *Radio Science*, **39**, RS1S05, doi:10.1029/2002RS002859.
- Hall, M. P. M., and Barclay, L. W., 1989. *Radiowave Propagation*. London: Peter Peregrinus.
- Hunsucker, R. D., 1991. *Radio Techniques for Probing the Ionosphere*. New York: Springer.
- Jakowski, N., Wilken, V., and Mayer, C., 2007. Space weather monitoring by GPS measurements on board CHAMP. *Space Weather*, **5**, S08006, doi:10.1029/2006SW000271.
- Khattatov, B., Murphy, M., Cruikshank, B., and Fuller-Rowell, T., (2004). Ionospheric corrections from a prototype operational

- assimilation and forecast system. In *Proceedings of IEEE Position, Location, and Navigation Symposium (PLANS)*. New York: Institute of Electrical and Electronic Engineers, pp. 518–526. Development of a physics-based reduced state Kalman filter for the ionosphere, *Radio Science*, **39**, RS1S04, doi:10.1029/2002RS002797.
- Komjathy, A., 1997. *Global Ionospheric Total Electron Content Mapping Using the Global Positioning System*. Ph.D. dissertation, Department of Geodesy and Geomatics Engineering Technical Report No. 188, University of New Brunswick, Fredericton, New Brunswick, p. 248.
- Komjathy, A., Sparks, L., Wilson, B., and Mannucci, A. J., 2005. Automated daily processing of more than 1000 ground-based GPS receivers to study intense ionospheric storms. *Radio Science*, **40**, RS6006, doi:10.1029/2005RS003279.
- Langley, R. B., 1996. Propagation of the GPS signals. In *GPS for Geodesy*, International School, Delft, 26 March–1 April, 1995. Springer, New York.
- Mandrake, L., Wilson, B., Wang, C., Hajj, G., Mannucci, A., and Pi, X., 2005. A performance evaluation of the operational Jet Propulsion Laboratory/University of Southern California Global Assimilation Ionospheric Model (JPL/USC GAIM). *Journal of Geophysical Research*, **110**, A12306, doi:10.1029/2005JA011170.
- Mannucci, A., Wilson, B., Yuan, D., Ho, C., Lindqwister, U., and Runge, T., 1998. A global mapping technique for GPS-derived ionospheric total electron content measurements. *Radio Science*, **33**(3), 565–582.
- Pi, X., Wang, C., Hajj, G. A., Rosen, G., Wilson, B. D., and Bailey, G. J., 2003. Estimation of ExB drift using a global assimilative ionospheric model: an observation system simulation experiment. *Journal of Geophysical Research*, **108**(A2), 1075, doi:10.1029/2001JA009235. Measurement, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. (lukas.mandrake@jpl.nasa.gov) C. Wang, Department of Mathematics, University of Southern California.
- Pi, X., Had, G. A., Wilson, B. D., Mannucci, A. J., Komjathy, A., Mandrake, L., Wang, C., and Rosen, I. G., 2004. 3-dimensional assimilative ionospheric modeling for regions of large TEC gradient, *Pmc ION NTM*, San Diego, January, 2004.
- Rius, A., Ruffini, G., and Cucurull, L., 1997. Improving the vertical resolution of ionospheric tomography with GPS occultations. *Geophysical Research Letters*, **24**(18), 2291–2294.
- Scherliess, L., Schunk, R. W., Sojka, J. J., Thompson, D. C., and Zhu, L., 2006. Utah state university global assimilation of ionospheric measurements Gauss-Markov Kalman filter model of the ionosphere: model description and validation. *Journal of Geophysical Research*, **111**, A11315, doi:10.1029/2006JA011712.
- Schunk, R. W., 2002. *Global Assimilation of Ionospheric Measurements (GAIM)*. Paper presented at ionospheric effects symposium, Office of Naval Res., Alexandria, VA.
- Schunk, R. W., and Nagy, A. F. (eds.), 2000. *Ionospheres: Physics, Plasma Physics, and Chemistry*. New York: Cambridge University Press, p. 333.
- Wang, C., Hajj, G., Pi, X., Rosen, I. G., and Wilson, B., 2004. Development of the global assimilative ionospheric model. *Radio Science*, **39**, RS1S06, doi:10.1029/2002RS002854.

Cross-references

Electromagnetic Theory and Wave Propagation
 Geodesy
 GPS, Occultation Systems
 Limb Sounding, Atmospheric
 Magnetic Field
 Radars

IRRIGATION MANAGEMENT

Steven R. Evett¹, Paul D. Colaizzi¹,
 Susan A. O'Shaughnessy¹, Douglas J. Hunsaker² and
 Robert G. Evans³

¹USDA-ARS Conservation and Production Research
 Laboratory, Bushland, TX, USA

²USDA-ARS, Maricopa, AZ, USA

³USDA-ARS, Sidney, MT, USA

Synonyms

Irrigation automation; Irrigation scheduling; Precision irrigation; Site-specific irrigation

Definition

Irrigation management. The decision process and action of applying a chosen depth of irrigation water using a chosen application method at a chosen time to achieve defined agronomic and economic objectives.

Supervisory control and data acquisition (SCADA) system. An electronic hardware and software system that acquires data from sensors and controls the operations of other hardware in a supervisory fashion.

Irrigation automation. The use of in situ, remotely sensed, or near-surface remotely sensed crop, soil, and micrometeorological properties sensed by a supervisory control and data acquisition system as inputs to a decision-making algorithm in the system, which then applies defined amounts of water at defined times automatically through control of an integrated irrigation application system.

Introduction

Irrigation has been practiced around the world for thousands of years. Irrigation works greater than 6,000 years old have been found in regions as diverse as Southeast Asia, the Indian subcontinent, Central Asia, the Middle East, Egypt, and North and South America. Some of these systems simply diverted water during floods and allowed it to spread on otherwise unflooded valley lands. Others were more sophisticated and allowed diversion of water from permanent stream flow so that irrigation could occur more or less on demand. Although irrigation management occurs anytime that irrigation occurs, only in the latter systems did management include the aspects of application timing and amount. Early irrigation management involved two methods of sensing irrigation needs, either probing of the soil or observation of plant characteristics such as wilting or folding leaves and leaf color. Unfortunately, changes in plant characteristics observable by the human eye tend to occur only after there has been some loss of potential yield. As irrigation methods grew more sophisticated, so did the means of deciding when and how much water to apply. Tubular probes for extracting deep soil

USDA is an equal opportunity provider and employer.

Robert G. Evans has retired.

cores allow the soil water content below the surface to be examined, and the “feel and appearance” method of judging soil water content was regularized through studies that established the relationships between the feel and appearance of a handful of soil and its particle size distribution and water potential energy. Practical instrumentation for soil water sensing for irrigation management became widespread after the introduction of the neutron moisture meter (NMM) in the 1950s, but it was not until the 1980s that automated soil water sensors based on electromagnetic (EM) physics became available and not until the 1990s that they became widely available. In the meantime, irrigation management based on electronic sensing of plant condition remained largely in the realm of scientific studies.

After the publication of Penman’s (1948) seminal paper, which established a theoretical basis for estimating plant water use (transpiration) from measurements of wind speed, solar radiation, air temperature, and humidity, a third and less direct paradigm for irrigation management was born. The new concept was based on the observation that crop water use varied with crop development (increasing with leaf area expansion) and with the evaporative demand, which could be estimated using Penman’s equation. Later, Jensen (1968) defined a crop coefficient, K_c , as the ratio of actual crop water use ET_c to a potential or reference water use, ET_o :

$$K_c = ET_c / ET_o \quad (1)$$

The variation of K_c with crop growth (and leaf area index) was empirically determined as a function of days after planting or, more recently, growing degree days (GDD), the latter providing somewhat more stable relationships for K_c (GDD). Knowing the function K_c (GDD) for a crop and region and possessing local weather data, local crop water use, ET , can be estimated as:

$$ET = K_c ET_o \quad (2)$$

This paradigm has become the dominant one in state and federal efforts to provide irrigation scheduling information to farmers in the USA as evidenced by networks of weather stations and corresponding water use prediction systems established by both federal agencies (e.g., the AgriMET program sponsored by the Bureau of Reclamation in the Pacific Northwest) and states (e.g., CIMIS in California, AZMET in Arizona, the TXHPET Network in Texas, and the Oklahoma MESONET). The best of these provide decision support to some degree. For example, for each major crop in the region, the TXHPET Network provides crop water use values for the last day, 3 and 7 days for the closest weather station, and usually for each of three or four planting dates that span the earliest and latest planting in the current season. For crops such as corn that have short-season and long-season varieties, separate sets of values are provided for each. This reduces the management time needed to make irrigation decisions to a simple calculation of water

already applied plus precipitation versus the crop water use. However, adoption by producers faces several obstacles, including lack of site specificity of weather stations, a need for corresponding in-field precipitation measurements, and an inability to characterize in-field variations in crop water use and water need. This is despite numerous and long-term efforts to improve the estimation of ET_o and K_c (GDD) by improving the theoretical underpinnings, taking into account plant cover effects and evaporation from soil more explicitly, and by improving siting and operation of weather stations (e.g., Monteith, 1965; Jensen et al., 1990; Allen et al., 1998; ASCE, 2005).

Given the current state of practice in irrigation management, the challenge for remote sensing is to provide irrigation management tools that allow producers to obtain comparable yield, yield quality, and water and nutrient use efficiencies profitably and sustainably, which means that such tools must be practical and cost-effective.

Spatial and temporal variability

Although traditional soil water, plant-based, and weather-based methods of irrigation scheduling address temporal variability of crop water need reasonably to very well, spatial variability is poorly addressed by these methods. Aerial and satellite imagery allow determination of spatial variability but often with temporal and spatial resolutions that are inadequate for day-to-day irrigation management (Jackson, 1984; Moran, 1994; Moran et al., 1997). The conjunctive use of coarse resolution thermal images with high-resolution images in the near-infrared and visible spectrums may provide solutions to the spatial resolution problem, but this is the subject of ongoing research (Kustas et al., 2004). These efforts are combined with those to merge data from satellites that return daily images with data useful for irrigation scheduling that are obtained only weekly or biweekly (Anderson et al., 2007). Such data fusion efforts were reviewed by Ha et al. (2013a, b). A key problem is that thermal infrared radiance data are the most useful for estimating crop ET , but these data are not available at high-enough temporal or spatial resolution and may even be lacking in new satellite platforms currently under development. The challenges that the status quo presents are being met by an array of near-surface remote sensing efforts, many of them using sensors that are mounted on moving irrigation systems (Evans and Sadler, 2008; Sadler et al., 2007) or on masts set in fields and some using aircraft platforms, including unmanned aerial vehicles (UAVs).

Remote sensing for irrigation management

Satellite and aerial remote sensing have become widespread only since the 1970s, although early efforts go back to the aerial photography from balloon platforms in the 1800s. As discussed elsewhere in this volume, the development of infrared film and electronic imaging in the visible spectrum and the invisible infrared and ultraviolet

spectrums allowed a new science of remote sensing of plant and soil condition to develop.

There are several approaches to irrigation management using remote sensing, including:

1. Scheduling irrigation to replace ET estimated from a reference ET (ET_o), calculated from local weather data, which is multiplied by a crop coefficient estimated with a crop coefficient function, $K_c(NDVI)$, where NDVI is the normalized difference vegetative index (NDVI) or a similar index adjusted for reflectance from soil. The NDVI is based on canopy irradiance in the red and near-infrared bands, which can be remotely sensed.
2. Scheduling irrigations at a fixed amount of water whenever a trigger to irrigate is generated by the crop water stress index (CWSI), which is estimated using remotely sensed T_s and local weather data.
3. Scheduling irrigations at a fixed amount when triggered by the time-temperature threshold index (TTTI) reaching a crop and region-specific value. The TTTI is calculated using T_s .
4. Scheduling irrigation to replace ET estimated with the field surface energy balance (FSEB), which uses remotely sensed surface temperature, T_s , determined from thermal infrared data, and data on canopy cover and surface emissivity inferred from the near-infrared (NIR) and visible bands.
5. Sensing of crop characteristics in order to guide timing, placement, and amount of fertilizer and water through irrigation (or fertigation) systems of various orders of precision. The characteristics, including crop cover fraction, nitrogen status of leaves, disease, and pest damage, all of which vary spatially and temporally, are inferred from various remotely sensed vegetative indices (VIs).

Of the five approaches listed, only the CWSI and the TTTI have been commercialized and used by irrigators, the latter recently under the name BIOTIC (The mention of trade names of commercial products in this entry is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.) (Upchurch et al., 1996) and the former since the 1980s. Practical use of both is quite limited currently. All of these approaches are the subjects of ongoing research efforts.

The ability to create broadband (thermal infrared to ultraviolet) images and to filter them into sub-band images led to breakthroughs in plant condition and cover sensing. Multispectral vegetation indices (VIs), such as the NDVI, are derived as ratios of signal strength in particular bands. Multispectral VIs have been widely researched as means to quantify various biophysical aspects of vegetation canopies, such as leaf area index (Moran et al., 1995), crop yield (Plant et al., 2000), and percent crop cover (Heilman et al., 1982). Thus, remote sensing of VIs provides a way to synoptically and instantaneously view crop conditions.

An approach that may improve the spatial representation of crop ET_c estimation is to incorporate remote sensing observations into irrigation scheduling protocols. A promising technique, introduced in the 1980s (Bausch and Neale, 1987), utilizes multispectral VIs to estimate corn crop coefficients. More recent research has shown that observations of multispectral VIs can provide real-time surrogates of crop coefficients for a variety of crops (Bausch, 1995; Neale et al., 2003; Hunsaker et al., 2005). Moreover, the use of remote sensing to infer the spatial distribution of K_c across the landscape can improve the ability of standard weather-based ET methods to more accurately estimate the spatial crop water use within an irrigated-field (Hunsaker et al., 2007) or at the farm-scale level (Johnson and Scholasch, 2005). The VI-based crop coefficient approach has strong practical appeal due to the longstanding familiarity and widespread use of crop coefficient methods and their relative operational simplicity. Also, this approach avoids some of the pitfalls with thermal bands, namely, the coarser spatial resolution and possible exclusion from future satellites. Nevertheless, implementation of the approach could be hindered by its reliance on empirical relationships between VIs and crop coefficients, by problems associated with the transferability of crop coefficient calibrations from one region to the next, and by timeliness and cost-effectiveness of the necessary imagery (Gowda et al., 2008).

Many methods of sensing plant water stress have been developed and studied, but only a few have impacted irrigation management. Dendrometers and leaf thickness and fruit size sensors have been used, mostly for high-value fruit and nut crops, but these approaches do not lend themselves to remote sensing. One of the earliest plant-based efforts that used near-surface remote sensing and that impacted irrigation management was the crop water stress index (CWSI) developed at the US Water Conservation Laboratory in the 1970s and 1980s (Idso et al., 1977, 1981; Jackson et al., 1981). The CWSI uses crop canopy temperature, typically sensed by an infrared thermometer, along with measured air temperature, humidity, wind speed, and solar radiation. The canopy-air temperature difference ($T_c - T_a$) is normalized to lower and upper limits of canopy-air temperature differences, which represent non-water-stressed and completely water-stressed crops, respectively. The CWSI is defined as

$$CWSI = \frac{(T_c - T_a)_m - (T_c - T_a)_l}{(T_c - T_a)_u - (T_c - T_a)_l} \quad (3)$$

where the subscript “m” denotes measured, “l” denotes a lower baseline (non-water-stressed crop), and “u” denotes an upper limit (completely water-stressed crop). The $(T_c - T_a)_l$ and $(T_c - T_a)_u$ limits may be estimated on an empirical (Idso et al., 1981) or theoretical (Jackson et al., 1981) basis. Both approaches require, at minimum, measurements of air temperature and humidity. The latter approach also requires wind speed, solar radiation, and

estimates of bulk canopy resistances. Meteorological parameters were not widely available until automated agricultural weather stations began to appear during the early 1980s (Howell et al., 1984). Furthermore, it was not until this time that basic data on bulk canopy resistances became established for a wide variety of crops and local conditions. Hence, the empirical CWSI received greater attention than its theoretical counterpart early on.

As initially developed in research and deployed for irrigation management, the CWSI employed a handheld infrared thermometer used to measure canopy temperatures at one time of day near solar noon. Although theoretically enticing, the CWSI has not been widely used for irrigation management due to various problems, including lack of a stable signal during cloudy periods, difficulty in determining the lower baseline and upper limit and nonuniqueness of these limits, and problems with measuring a meaningful canopy temperature that is representative of the field crop when soil background is visible, which might occur during part or all of the irrigation season. Moran et al. (1994) extended the CWSI concept where the upper and lower limits were defined as mixtures of soil and vegetation temperatures and termed this the water deficit index (WDI). However, both the CWSI and WDI may be sensitive to small errors in the upper and lower temperature limits, which often results in values outside the theoretical range from zero to one. The effect of sunlit vs. shaded soil and a well-watered crop underlain by dry soil (e.g., subsurface drip irrigated soil, Colaizzi et al., 2003) may further confound errors in estimating these limits.

Although the early work with the CWSI was done using handheld infrared thermometers to sense the thermal infrared radiation from crop canopies, later work has shown that maps of the CWSI or WDI can be created from aerial and satellite images or images compiled from ground-based sensors aboard self-propelled irrigation systems (Colaizzi et al., 2003; Peters and Evett, 2008; O'Shaughnessy et al., 2008). For example, an empirical CWSI (eCWSI) was determined to be a robust stress index for irrigated vineyards:

$$\text{eCWSI} = \frac{T_c - T_w}{T_{\text{dry}} - T_w} \quad (4)$$

where T_c was the temperature ($^{\circ}\text{C}$) of the center of the crop canopy computed from aerial thermal images, T_w was the temperature from an artificial wet surface that acted as a substitute for the well-watered baseline temperature, and T_{dry} was estimated by adding 5°C to the maximum dry bulb temperature recorded for the specific field day (Jones, 1992; Möller et al., 2007). Similar relationships were shown for cotton grown in the Texas High Plains where calculated eCWSI values were compared to stem leaf water potential (Ψ_L) and in the South of Israel where comparisons were made with Ψ_L and stomatal conductance (Evett and Alchanatis, 2007). These empirical relationships were more robust than the theoretical model of

Jackson et al. (1981) relating crop canopy temperatures to crop water status, and they provided a basis for spatial mapping of Ψ_L at the field scale. In areas where irrigation decisions are based on Ψ_L , this modeling may replace the labor-intensive efforts of Ψ_L measurements.

Although fundamentally related to plant water status, the CWSI faces a number of problems that have so far prevented its widespread use. The well-watered baseline is a function of plant species and for a single species can vary by important amounts between varieties (e.g., cotton: Hatfield et al., 1987). Empirical determinations of lower baseline vary depending on how "well-watered" is defined. For partial canopies, emittance from soil can bias the measure of T_c , a problem for all methods that use T_c . Also, viewing angle of the infrared sensor will affect the temperature reading, as will flowering and seed filling in some crops (corn, sorghum, and wheat). Taking readings at an angle from nadir and taking readings from opposing viewpoints are techniques used to reduce these biases. However, these techniques are not easily achieved from satellite and aircraft platforms. For these reasons, the CWSI in its many forms has not been widely adopted for use in irrigation management.

The TTTI is the number of minutes in a day that a canopy is above a species-specific threshold canopy temperature. If the TTTI exceeds a threshold time (min), which is specific to a species and climatic region, then an irrigation is triggered. The threshold time was determined by finding the average number of minutes in a day throughout the irrigation season that the canopy of a well-watered crop was above the threshold temperature. The threshold temperature was originally conceived as the temperature at which photosynthetic assimilation peaks (Burke and Oliver, 1993). However, Evett et al. (1996, 2000, 2006a) showed that the threshold temperature could be varied as long as a corresponding threshold time was determined from canopy temperatures measured on a well-watered crop. The method suffers from the same viewing angle problems that affect the CWSI. And the threshold temperature may be variety specific, not just species specific. Also, the method requires a continuous record of canopy temperature throughout the daylight hours, something not readily achieved using airborne or satellite platforms. The latter problem was overcome by Peters and Evett (2004) who devised an algorithm to scale a one-time-of-day canopy temperature measurement so as to estimate a continuous course of canopy temperatures for daytime, except for 1 h after sunrise and 1 h before sunset, with sufficient accuracy to trigger irrigations using the TTTI ($<1^{\circ}\text{C}$ error). This algorithm may have utility in scaling thermal image data from airborne or satellite platforms to produce maps of TTTI values and thus maps of crop water stress over large areas.

Thermal radiometric measurements of crop canopy temperature for automatic irrigation scheduling using the TTTI have been successful for drip-irrigated plots (Evett et al., 1996) and low-energy precision application (LEPA) irrigation using a center pivot sprinkler (Peters and

Evett, 2008). This simplified approach avoids the complexity of the CWSI by relying on the feedback effect of irrigation, which by providing water to plants allows increased transpiration and resultant cooling of the canopy. Wired thermal radiometers are now being replaced with wireless infrared thermometers and wireless sensor networks to establish remote links for data collection, communication, and control of irrigation systems (O'Shaughnessy and Evett, 2007). The specific implementation of the TTTI embodied in the BIOTIC patent has been licensed and commercialized since 2008. Although it is difficult to envision application of the TTTI using data from satellite or aircraft platforms, it is readily applied using near-surface remote sensing with sensors mounted on towers or moving irrigation system superstructures. In fact, use of moving irrigation systems as sensing platforms has allowed the implementation of field mapping of crop water stress levels and relative yield potential (Peters and Evett, 2007) as well as irrigation control.

The field surface energy balance (FSEB) has long been recognized for its promise in calculating the latent heat flux, LE , due to ET , from the other energy balance terms in

$$0 = LE + G + H + R_n \quad (5)$$

where G is soil heat flux, H is sensible heat flux, and R_n is net radiation (all fluxes taken positive toward the surface, often in $W\ m^{-2}$). In field research, many or all of these fluxes are measured, but from a remote sensing perspective, the surface brightness in the thermal infrared (used to estimate T_s) and the brightness in the near-infrared and visible bands are used to estimate R_n and H . The value of G is taken as a fraction of R_n or some function of R_n and plant cover (estimated from a VI), and ET as a depth of water per unit time ($m\ s^{-1}$, positive toward the atmosphere) is evaluated as the residual

$$ET = -LE/(\lambda\rho_w) = (R_n + G + H)/(\lambda\rho_w) \quad (6)$$

where $H = \rho_a c_p (T_s - T_a)/r_a$; T_s is assumed equal to the aerodynamic surface temperature, T_o ($^{\circ}C$); r_a is aerodynamic resistance, which is calculated as a function of crop height, h_c , and wind speed, u_z ($m\ s^{-1}$), measured at an elevation z (m); c_p is the heat capacity of air ($\sim 1,003\ kJ\ kg^{-1}\ K^{-1}$); λ is the latent heat of vaporization ($\sim 2.45 \times 10^6\ J\ kg^{-1}$); ρ_w is the density of water ($\sim 1,000\ kg\ m^{-3}$); and ρ_a is air density ($kg\ m^{-3}$, $\rho_a \approx 1.291 - 0.00418T_a$). When R_n , G , T_s , T_a , h_c , and u_z are measured in the field, Equation 6 can estimate ET with fair accuracy (Kimball et al., 1999) for full-cover crop surfaces. Because energy balance models have more numerous applications besides cropped surfaces or irrigation management, they have been investigated more extensively compared with the VI-based crop coefficient approach.

Remote sensing approaches to FSEB evaluation for ET are numerous and include both single-surface approaches that assume that the crop can be modeled as a big leaf (i.e., Equation 6) (e.g., Bastiaanssen et al.,

2005; Allen et al., 2007a, b) and two-surface approaches that evaluate the energy and water balances for both canopy and soil surfaces and thus must estimate the crop cover fraction (e.g., Evett and Lascano, 1993; Norman et al., 1995; Kustas and Norman, 1999; Colaizzi et al., 2012). While providing useful knowledge of regional ET and its spatial and temporal distribution, FSEB predictions that rely on satellite images are not used for irrigation management due to lack of daily data with sufficient resolution to match the field scales at which irrigation is managed (Gowda et al., 2008). Attempts to resolve this problem with existing satellite data involve using infrequent, higher-resolution data such as that from Landsat (90 m resolution in the thermal IR, 16 day repeat time) in combination with lower-resolution more frequent data such as that from MODIS (1,000 m resolution in the thermal IR, daily). Such a combination was demonstrated with the DisALEXI (disaggregated atmosphere land exchange inverse) model (Anderson et al., 2007; Kustas et al., 2007; Norman et al., 2003), but no testing of this approach for irrigation management has ensued. While aircraft platforms could resolve some of these problems, so far there has not been widespread use of aircraft platforms to provide imaging for irrigation management, probably due to cost.

The FSEB using remotely sensed data typically provides an instantaneous value of ET , which must be scaled to daily ET using various methods such as the evaporative fraction (which is assumed constant during daylight hours) or reference ET (which is also assumed constant relative to latent heat flux during daylight hours) (Colaizzi et al., 2006). Models of the FSEB that use remote sensing include the two-source model (TSM, Norman et al., 1995; Kustas and Norman, 1999), the surface energy balance algorithm for land (SEBAL, Bastiaanssen et al., 1998), and mapping ET with internalized calibration (METRIC, Allen et al., 2007a, b). Gowda et al. (2008) reviewed these and several other approaches and reported that daily ET estimation errors ranged from 3 % to 35 % but that almost all studies compared FSEB ET with ET sensed by Bowen ratio or eddy covariance methods, which themselves have large errors. Colaizzi et al. (2012) evaluated the two-source energy balance model of Norman et al. (1995) and Kustas and Norman (1999) where crop ET was measured with weighing lysimeters at Bushland, Texas, and radiometric temperature was measured with infrared thermometers that viewed the lysimeter surface. Crops included fully irrigated corn, alfalfa, soybean, and winter wheat; deficit-irrigated cotton; and dryland grain sorghum. This study differed from previous two-source energy balance model testing studies, which mainly used Bowen ratio or eddy covariance systems to indirectly estimate ET and used remotely sensed data from satellite or airborne platforms to run the models. The root mean squared errors between modeled and observed energy fluxes were slightly greater in the Colaizzi et al. (2012) study compared with the previous studies, which was probably related to the

methods used to estimate ET and radiometric surface temperature.

The vast majority of energy balance model testing studies have been at the watershed or regional scale and used satellite data, which are too coarse and/or infrequent for irrigation management. Moreover, ET estimates from regional models are often inaccurate due to problems with correctly estimating the surface radiation balance components (Berbery et al., 1999). However, several studies, using either ground-based or airborne sensor platforms to achieve suitable spatial scales (i.e., a few meters or less), have shown the utility of energy balance and stress index models for irrigation management (Colaizzi et al., 2003; French et al., 2007), including automatic irrigation scheduling and control (Evelt et al., 1996, 2000, 2006a; Peters and Evelt, 2008). One outcome of the years of research on remote sensing using satellite platforms has been the cross-fertilization of the irrigation engineering field where ground-based sensor platforms are more likely to be utilized but where the principles of remote sensing are just as applicable. Ground-based or near-surface remote sensing provides data of high temporal and spatial resolution that can be very useful in irrigation management, but largely without the need for correction for interference from the atmosphere. Remote sensing principles and techniques are routinely used, including spectral data acquisition and processing, VI relationships to plant cover and condition, thermal infrared data acquisition, relationships to plant condition, and energy balance.

In addition to the main avenues of research and application involving remote sensing for irrigation management, there are some areas that have not seen serious or successful efforts due to either intractable problems, newness of the approach, or lack of direct applicability. Several researchers have documented relationships between surface soil water content or soil profile water content and temperature indices. Sadler et al. (2000) found that the canopy-air temperature difference, ($T_c - T_a$), explained 60 % of the variation of the fraction of available water in the top meter of soil. Colaizzi et al. (2003) developed relationships between the CWSI of flood-irrigated cotton on a sandy loam with a soil water stress index and with both the fractional soil water content depletion and depth of depletion. However, they stressed that the relationship was strongest for severely depleted soil and water-stressed plants and would lose relevance for frequently and more adequately irrigated crops. Passive and active microwave radiation and synthetic aperture radar have been used for surface water content estimation (top few centimeters of soil) but encounter the problems of imprecisely known sensing depth, interference from vegetation, and spatial resolution too coarse for irrigation management. Characterization of soil roughness needed to produce accurate soil water content estimates is not well understood, and only regional-scale mapping is possible (Verhoest et al., 2008). A series of field campaigns with various types of ground truthing have explored this and related approaches

to soil water content sensing (Washita'92, Washita'94, SGP97, SGP99, SMEX02, SMEX03, SMEX04, SMEX05, CLASIC – Land, SMAPVEX08; e.g., Jackson et al., 1995, 1999, 2005, 2008; Jackson and Hsu, 2001). This approach is more useful for synoptic studies than for irrigation management (e.g., Zhan et al., 2008). Some recent work has focused on using thermal remote sensing surface temperature data as input to a FSEB model with a soil water component in order to drive estimates of profile water content (Crow et al., 2008). This is similar to attempts made using near-surface remote sensing of canopy temperatures to drive models of the soil-plant-atmosphere continuum. To date, these approaches have not been useful in field-scale irrigation management.

Current obstacles

Remotely sensed data must meet several criteria to be suitable for irrigation management, but present satellite and airborne platforms do not meet one or more of these requirements, which include spatial resolution, repeat frequency (i.e., temporal resolution), accuracy, and turnaround time. Jackson (1984) discussed these requirements and concluded that spatial resolution must be on the order of a few meters, repeat frequency no more than a few days (although hourly would be most desirable), and turnaround time (i.e., the interval from actual field measurements to the usable data product) no more than a few minutes. Since then, irrigated area under moving and drip irrigation systems has greatly increased (e.g., >80 % in the Southern Great Plains), and these systems require data at intervals not greater than a day. Present satellite platforms with thermal bands have a trade-off between spatial and temporal resolution. Although satellite data have been successfully used to construct spatially distributed ET maps at the watershed (Allen et al., 2007b) or continental scale (Kustas et al., 2004), they lack adequate spatial resolution for irrigation management and have turnaround times on the order of weeks or more due to extensive processing requirements (i.e., atmospheric and geometric correction). Sharpening algorithms, such as TsHARP, do not accurately recover information at sub-field scales suitable for irrigation management (Agam et al., 2007, 2008). Airborne platforms can potentially meet both the spatial and temporal requirements but are cost prohibitive and also suffer from undue turnaround times because automated processing algorithms have yet to be developed. Moran (1994) demonstrated the shortcomings of both satellites and aircraft for irrigation management, which were mainly due to turnaround time, but also due to inexplicable lapses in data delivery from cooperating entities. More recently, there have been efforts to adopt unmanned air vehicles (UAVs) to carry sensors suitable for on-farm management applications (Herwitz et al., 2004; Xiang and Tian, 2007; Chao et al., 2008). Unresolved issues include the need to reduce sensor package mass, reliable radio control, lack of control

in windy conditions, orthogonal correction of non-nadir views, and geo-referencing.

Rigorous FSEB model assessment and testing is contingent on both accurate ground truth and remotely sensed data. High-quality ground truth data are particularly expensive and difficult to obtain, and the simultaneous acquisition of remotely sensed data is indeed challenging. These factors have impeded development of more robust models. Weighing lysimeters are the most direct and potentially accurate method of ET measurement (Howell et al., 1995); however, they are expensive, not portable, and require highly trained personnel to operate and maintain. ET may also be derived with good accuracy as the residual of the soil water balance if measurements of irrigation, rainfall, and soil water throughout the profile are available, and provided other sources and sinks that are difficult to measure can be minimized or eliminated (e.g., run-on, runoff, deep percolation, and upward capillary flux from a shallow water table). However, accurate measurement of the soil profile water content has been limited to destructive and labor-intensive gravimetric sampling or the neutron moisture meter (NMM), which is also labor intensive and cannot be operated unattended. Electromagnetic soil water sensors presently on the market, which are appealing in that they may be automated, suffer from very poor accuracy in most soils (Evetts et al., 2006b, 2008, 2009; Mazahreh et al., 2008). Other techniques of estimating ET include Bowen ratio, eddy covariance, and meteorological flux towers. These systems are portable, relatively low cost, and easily automated, and hence have been the most popular ground truth techniques for model testing (validation) studies. However, the Bowen ratio method assumes that the conductivity for water vapor flux, K_v , equals the conductivity for heat flux, K_h , and is therefore subject to errors during advective conditions and can also be sensitive to fetch and instrument bias (Todd et al., 2000). Eddy covariance techniques sense turbulent fluxes across a field directly, but the energy balance is often not closed (Evetts et al., 2012a; Twine et al., 2000) even when additional procedures or corrections are used. Meteorological flux towers have been used extensively in desert environments (Kustas and Norman, 1999), but their accuracies over cropped surfaces or in advective environments are presently not known. At present, the most accurate ground truth is from weighing lysimeters and soil water balance ET estimates made using the NMM. Unfortunately, most FSEB model validation efforts have not been able to use those methods. A recent example of using weighing lysimeter and soil water balance measures of ET is, however, available in a special issue of *Advances in Water Resources* (Evetts et al., 2012b).

Ground-based sensors aboard farm machinery, such as self-propelled irrigation systems, are an alternative to satellite or airborne platforms. In regions where center pivot or lateral-move irrigation systems are popular, they appear to be a logical sensor platform for irrigation management, since they pass over the field at regular

intervals. Ground-based sensors may greatly reduce turnaround time because they do not require extensive processing such as atmospheric or geometric correction. Phene et al. (1985) is an early example of deployment of infrared thermometers aboard a self-propelled irrigation system. However, thermal radiometric sensors are responsive to soil emittance and to reflection from vegetative canopy: Less than full canopy cover may cause false-positive irrigation triggers in the early growing season and thermal measurements from mixed pixels of sunlit soil and vegetation can provide unduly high temperature readings. Discrimination between thermal radiance from soil and vegetation in low cover or leaf area index situations is a problem still under active investigation, and solution will probably require imaginative combination of multispectral data, sensor view angles, and understanding of canopy and soil characteristics.

Center pivot and linear move irrigation systems generally apply water quite uniformly; however, substantial variations in soil properties and water availability exist across most fields. In these cases, the ability to manage site-specific irrigation to match spatially and temporally variable conditions can offer opportunities for increased application efficiencies, reduced environmental impacts, more effective agrichemical use, and the potential to improve crop yields and quality. In addition, these types of systems afford the opportunity to precisely manage deficit irrigation strategies. The development of in-field sensor-based control of site-specific applications of water and water-soluble nutrients through the irrigation system offers an effective means to implement site-specific technologies, but the seamless integration of sensors, irrigation control, data interface, software design, and communication at costs that are in balance with the profit advantages of site-specific applications is challenging.

Several researchers have investigated the potential use of a feedback from wireless in-field sensing systems to control variable-rate irrigation systems, but few have fully integrated these systems. An exception is the TTTI system implemented for automatic irrigation control of drip and center pivot irrigation (Evetts et al., 2006a; Peters and Evetts, 2008). Miranda et al. (2003) used a closed-loop irrigation system and determined irrigation amount based on distributed soil water measurements. Shock et al. (1999) used radio transmission for soil moisture data from data loggers to a central computer logging site. Wall and King (2004) explored designs for smart soil moisture sensors and sprinkler valve controllers to implement plug-and-play technology and proposed architectures of distributed sensor networks for site-specific irrigation automation (King et al., 2000). Kim et al. (2008, 2009) used distributed sensor networks and GPS with Bluetooth® wireless communications to control water applications with off-site computers. Software design for automated irrigation control has been studied by Abreu and Pereira (2002). They designed and simulated solid set sprinkler irrigation systems by using software that allowed the design of a simplified layout of the irrigation system.

The coordination of control with data from sensors is effectively managed using data networks and low-cost microcontrollers (Wall and King, 2004). A hardwired system from in-field sensing station to a base station requires extensive time and cost to install and maintain. It may not be feasible to get the system hard wired for long distances, and it may not be acceptable to growers because it can interfere with normal farming operations and the maintenance costs may be unacceptable. A wireless data communication system can provide dynamic mobility and cost-free relocation. Radio frequency technology has been widely adopted in consumer wireless communication products, and it provides numerous opportunities to use wireless signal communication in agricultural systems. Industrial wireless standards such as the ZigBee protocol are open standards that allow integration of sensors and equipment from different manufacturers into a supervisory control and data acquisition (SCADA) system (O'Shaughnessy and Evett, 2007, 2008). Present challenges include meeting power requirements of remote sensors, radio interference, cost reduction, interfacing with existing irrigation control equipment, and development of rugged and inexpensive but accurate sensors (e.g., reflectance photodiodes and infrared thermometers).

Importance of unified technologies

Inherent in the evolution of on-farm water management is the integration of irrigation, fertilizer, and pest management strategies into systems that optimize total management practices for temporal and spatial variability using precision agriculture tools. This integration will require broad systems approaches that physically and biologically optimize irrigations with respect to water delivery and application schemes, rainfall, critical growth stages, soil fertility, location, and weather. Spatial and temporal management strategies will need to address site-specific crop requirements for water, nutrition, pest management, and irrigation scheduling in both agrarian and urban settings.

Improved irrigation technologies, while usually requiring less field labor, often demand more intensive management. Thus, support aids must also be developed that improve the producer's ability to implement decisions quickly and easily; however, this also requires control and monitoring systems. Decision support systems are needed to help make timely decisions based on complex inputs. Climatic variations and pest outbreaks require precisely timed and placed water and chemical applications on a daily and seasonal basis. The decision support process must also provide accurate predictions of crop water use, application efficiencies, and uniformities to improve management flexibility.

The commercialization of irrigation automation is dependent on the development of economical and reliable wireless sensors and sensor networks. Communication modules for each sensor type and network must be interchangeable. The progress of precision irrigation depends

on the seamless automatic operation of irrigation systems accomplished by the flexible integration of controllers and sensors, algorithms for data acquisition and control, compatible sensor hardware that is field installable and easy to maintain (plug-and-play format), and user-friendly interfaces that provide practical, robust, and timely decision-making support to the grower.

Conclusions

Ensuring the success of irrigated farming enterprises will require the development of reliable, timely information on field and plant status to support decision making. Remote sensing of plant and soil status using integrated satellite, aerial, and field-level plant and soil-based sensor systems can provide information on plant and soil nutrient and water status. Nevertheless, this technology needs further development to improve spatiotemporal modeling and hands-on use for on-farm management as well as irrigation district operations. Better systems and methods capable of measuring specific plant parameters (e.g., nutrient status, water status, disease, and competing weeds) on a timely basis are becoming available and are expected to provide information required to enhance use in crop modeling and improve within-season management.

Several remote sensing approaches have been shown to provide spatially distributed information on ET and plant water stress that is meaningful for irrigation management. These approaches use reflectance and thermal radiance measurements of the surface; those discussed herein included water stress indices, vegetation index (VI)-based crop coefficients, and surface energy balance models. Water stress indices are generally used to detect plant water stress and may be useful for appropriately timing irrigation events. The VI-based crop coefficient approach can be used to estimate crop ET when a reference ET value is available. This approach is essentially the same as the widely used crop coefficient-reference ET paradigm, where crop coefficients are based on calendar days or heat units, but VI-based crop coefficients estimated from images provide spatial information of actual crop conditions, unlike conventional crop coefficients. Surface energy balance models are generally classed as either single-layer or two-layer approaches, where the latter computes the energy fluxes of the soil and vegetation separately. Both types of energy balance models compute instantaneous latent heat flux, which can be converted and scaled to daily ET. The VI-based crop coefficient approach requires reflectance measurements but not thermal measurements as do water stress indices and energy balance models. However, the VI-crop coefficient approach is more empirically based and subject to local conditions, which may reduce geographic transferability. Energy balance models have been researched more extensively than VI-based crop coefficients, as they are applicable to other vegetated surfaces besides crops.

Routine application of remote sensing to irrigation management remains subject to several obstacles, which are generally related to model robustness and sensor platform cost, timeliness, and spatial resolution. The development of robust models has been hampered somewhat by inherent challenges in measuring ET or plant water stress for ground truth while simultaneously measuring surface reflectance and temperatures. The need for accurate and robust models is critical if they are to provide meaningful decision support for the rational allocation and timing of irrigation water. Just as critical is a fundamental understanding of model limitations, so that misapplication may be avoided. Sensor platforms, in order to be suitable for irrigation management, must provide data at spatial resolutions of a few meters, at least daily repeat frequencies, and have turnaround times of a few minutes. Existing satellite and airborne platforms do not meet one or more of these criteria. Key satellite platforms for thermal infrared data with sufficient spatial (but not temporal) resolution for irrigation management (LandSat 5 and 7) are nearing the end of their useful lives, and there is no near-term replacement, which makes expediting of the only new satellite planned with thermal IR capability (HypIRI, with 45 m resolution and a 5 day cycle) essential (Anderson and Kustas, 2008). Unmanned air vehicles (UAVs) and ground-based sensors aboard self-propelled irrigation systems have the potential to avoid some of the pitfalls of satellite or airborne platforms but nonetheless have other technical and cost issues to be overcome. At a time when thermal infrared data from satellite remote sensing is finally being used for effective water resource management at large scales (Allen et al., 2008), and when the technology for managing irrigation at small scales is becoming well developed and would be applicable if resolution in time and space were improved, the USA should be highly motivated to, rather than end its efforts, become once again a leader in space-based platforms for thermal infrared imaging at high spatial and temporal resolution.

Irrigated agriculture will be expected to provide more crops for a growing world population with less water resources. Also, farm consolidation continues in many regions, resulting in greater area being managed by fewer personnel. The need for better irrigation management tools that reduce management time is apparent. Remote sensing, with its capabilities of spatially distributed information, is one technology that may fulfill this need. Therefore, research and development efforts should continue to emphasize model validation and refinement and more suitable platforms and sensors capable of delivering timely irrigation management guidance and control at scales small enough to affect within-field management.

Bibliography

Abreu, V. M., and Pereira, L. S., 2002. *Sprinkler Irrigation Systems Design Using ISAMim*. ASABE Paper No. 022254. St. Joseph: ASABE.

- Agam, N., Kustas, W. P., Anderson, M. C., Li, F., and Colaizzi, P. D., 2007. Utility of thermal sharpening over Texas high plains irrigated agricultural fields. *Journal of Geophysical Research*, **112**, D19110, doi:10.1029/2007JD008407.
- Agam, N., Kustas, W. P., Anderson, M. C., Li, F., and Colaizzi, P. D., 2008. Utility of thermal image sharpening for monitoring field-scale evapotranspiration over rainfed and irrigated agricultural regions. *Geophysical Research Letters*, **35**, L02402, doi: 10.1029/2007GL032195.
- Allen, R. G., Periera, L. S., Raes, D., and Smith, M., 1998. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*. Rome: United Nations FAO. Irrig. and Drainage Paper No. 56.
- Allen, R. G., Tasumi, M., and Trezza, R., 2007a. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)-model. *Journal of Irrigation and Drainage Engineering*, **133**(4), 380–394.
- Allen, R. G., Tasumi, M., Morse, A., Trezza, R., Wright, J. L., Bastiaanssen, W., Krambler, W., Lorite, I., and Robinson, C. W., 2007b. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) – applications. *Journal of Irrigation and Drainage Engineering*, **133**(4), 395–406.
- Allen, R. G., Toll, D., Kustas, W., and Kleissl, J., 2008. From high overhead: ET measurement via remote sensing. *Southwest Hydrology*, **7**(1), 30–32.
- Anderson, M., and Kustas, W., 2008. Thermal remote sensing of drought and evapotranspiration. *EOS Transaction, American Geophysical Union*, **89**(26), 233–234.
- Anderson, C. M., Kustas, W. P., and Norman, J. M., 2007. Upscaling flux observations from local to continental scales using thermal remote sensing. *Agronomy Journal*, **99**, 240–254.
- ASCE, 2005. In Allen, R. G., Walter, I. A., Elliot, R. L., Howell, T. A., Itenfisu, D., Jensen, M. E., and Snyder, R. L. (eds.), *The ASCE Standardized Reference Evapotranspiration Equation*. Reston, VA: American Society of Civil Engineers.
- Bastiaanssen, W. G. M., Menenti, M., Feddes, R. A., and Holtslang, A. A., 1998. A remote sensing surface energy balance algorithm for land (SEBAL): 1 formulation. *Journal of Hydrology*, **212–213**, 198–212.
- Bastiaanssen, W. G. M., Noordman, E. J. M., Pelgrum, H., Davids, G., Thoreson, B. P., and Allen, R. G., 2005. SEBAL model with remotely sensed data to improve water-resources management under actual field conditions. *Journal of Irrigation and Drainage Engineering*, **131**(1), 85–93.
- Bausch, W. C., 1995. Remote sensing of crop coefficients for improving the irrigation scheduling of corn. *Agricultural Water Management*, **27**(1), 55–68.
- Bausch, W. C., and Neale, C. M. U., 1987. Crop coefficients derived from reflected canopy radiation: a concept. *Transactions of ASAE*, **30**(3), 703–709.
- Berbery, E. H., Mitchell, K. E., Benjamin, S., Smirnova, T., Ritchie, H., Hogue, R., and Radeva, E., 1999. Assessment of land-surface energy budgets from regional and global models. *Journal of Geophysical Research*, **104**(D16), 19329–19348.
- Burke, J. J., and Oliver, M. J., 1993. Optimal thermal environments for plant metabolic processes (*Cucumis sativus* L.) (light-harvesting chlorophyll a/b pigment-protein complex of photosystem II and seedling establishment in cucumber). *Plant Physiology*, **102**(1), 295–302.
- Chao, H., Baumann, M., Jensen, A., Chen, Y., Cao, Y., Ren, W., and McKee, M., 2008. Band-reconfigurable multi-UAV-based cooperative remote sensing for real-time water management and distributed irrigation control. In *Proceedings of the 17th World Congress, International Federation. Automatic Control*, July 6–11, 2008, Seoul, pp. 11744–11749.

- Cohen, Y., Alchanatis, V., Meron, M., Saranga, Y., and Tsipris, J., 2005. Estimation of leaf water potential by thermal imagery and spatial analysis. *Journal of Experimental Botany*, **56**(417), 1843–1852.
- Colaizzi, P. D., Barnes, E. M., Clarke, T. R., Choi, C. Y., Waller, P. M., Haberland, J., and Kostrzewski, M., 2003. Water stress detection under high frequency sprinkler irrigation with water deficit index. *Journal of Irrigation and Drainage Engineering*, **129**(1), 36–43.
- Colaizzi, P. D., Evett, S. R., Howell, T. A., and Tolk, J. A., 2006. Comparison of five models to scale daily evapotranspiration from one time of day measurements. *Transactions of the ASABE*, **49**(5), 1409–1417.
- Colaizzi, P. D., Bliessner, R. D., and Hardy, L. A., 2008. A review of evolving critical priorities for irrigated agriculture. In *Proceeding of the World Environmental and Water Resources Congress 2008 Ahupua'a*. May 12–16, 2008, Honolulu: American Society of Civil Engineers, Environmental and Water Resources Institute.
- Colaizzi, P. D., Evett, S. R., Howell, T. A., Gowda, P. H., O'Shaughnessy, S. A., Tolk, J. A., Kustas, W. P., and Anderson, M. C., 2012. Two source energy balance model-refinements and lysimeter tests in the Southern High Plains. *Transactions of the ASABE*, **55**(2), 551–562.
- Crow, W. T., Kustas, W. P., and Prueger, J., 2008. Monitoring root-zone soil moisture through the assimilation of a thermal remote sensing-based soil moisture proxy into a water balance model. *Remote Sensing of Environment*, **112**, 1268–1281.
- Evans, R. G., and Sadler, E. J., 2008. Methods and technologies to improve efficiency of water use. *Water Resources Research*, **44**, W00e04, doi: 10.1029/2007WR006200.
- Evett, S. R., 2007. Soil water and monitoring technology. In Lascano, R. J., and Sojka, R. E. (eds.), *Irrigation of Agricultural Crops*, 2nd edn. Madison, WI: American Society of Agronomy/Crop Science Society of America/Soil Science Society of America. Agronomy Monogr. Vol. 30, pp. 25–84.
- Evett, S. R., and Alchanatis, V. L., 2007. *Improved analysis of thermally sensed crop water status and mapping spatial variability for site specific irrigation scheduling*. Final Report to BARD and the Texas Department of Agriculture on project TIE04-01. USDA-ARS Conservation and Production Research Laboratory, Bushland.
- Evett, S. R., and Lascano, R. J., 1993. ENWATBAL.BAS: a mechanistic evapotranspiration model written in compiled BASIC. *Agronomy Journal*, **85**(3), 763–772.
- Evett, S. R., Howell, T. A., Schneider, A. D., Upchurch, D. R., and Wanjura, D. F., 1996. Canopy temperature based automatic irrigation control. In Camp, C. R., et al. (eds.), In *Proceedings of the International Conference Evapotranspiration and Irrigation Scheduling*, San Antonio, ASAE, St. Joseph, November 3–6, 1996, pp. 207–213.
- Evett, S. R., Howell, T. A., Schneider, A. D., Upchurch, D. R., and Wanjura, D. F., 2000. Automatic drip irrigation of corn and soybean. In Evans, R. G., Benham, B. L., and Trooien, T. P. (eds.), In *Proceedings of the 4th Decennial National Irrigation Symposium*. Phoenix, November 14–16, pp. 401–408.
- Evett, S. R., Peters, R. T., and Howell, T. A., 2006a. Controlling water use efficiency with irrigation automation: cases from drip and center pivot irrigation of corn and soybean. In *Proceedings of the 28th Annual Southern Conservation Systems Conference*, Amarillo, June 26–28, 2006, pp. 57–66.
- Evett, S. R., Tolk, J. A., and Howell, T. A., 2006b. Soil profile water content determination: sensor accuracy, axial response, calibration, temperature dependence and precision. *Vadose Zone Journal*, **5**, 894–907.
- Evett, S. R., Heng, L. K., Moutonnet, P., and Nguyen, M. L., (eds.). 2008. *Field Estimation of Soil Water Content: A Practical Guide to Methods, Instrumentation, and Sensor Technology*. IAEA-TCS-30. Vienna: International Atomic Energy Agency, ISSN 1018–5518.
- Evett, S. R., Tolk, J. A., and Howell, T. A., 2009. Soil profile water content determination in the field: spatio-temporal variability of electromagnetic and neutron probe sensors in access tubes. *Vadose Zone Journal*, **8**, 926.
- Evett, S. R., Schwartz, R. C., Howell, T. A., Baumhardt, R. L., and Copeland, K. S., 2012a. Can weighing lysimeter ET represent surrounding field ET well enough to test flux station measurements of daily and sub-daily ET? *Advances in Water Resources*, **50**, 79–90.
- Evett, S. R., Kustas, W. P., Gowda, P. H., Prueger, J. H., and Howell, T. A., 2012b. Overview of the Bushland Evapotranspiration and Agricultural Remote sensing EXperiment 2008 (BEAREX08): a field experiment evaluating methods quantifying ET at multiple scales. *Advances in Water Resources*, **50**, 4–19.
- French, A. N., Hunsaker, D. J., Clarke, T. R., Fitzgerald, G. J., Luckett, W. E., and Pinter, P. J., Jr., 2007. Energy balance estimation of evapotranspiration for wheat grown under variable management practices in Central Arizona. *Transactions of the ASABE*, **50**(6), 2059–2071.
- Gowda, P. H., Chavez, J. L., Colaizzi, P. D., Evett, S. R., Howell, T. A., and Tolk, J. A., 2007. Remote sensing based energy balance algorithms for mapping ET: current status and future challenges. *Transactions of the ASABE*, **50**(5), 1639–1644.
- Gowda, P. H., Chavez, J. L., Colaizzi, P. D., Evett, S. R., Howell, T. A., and Tolk, J. A., 2008. ET mapping for agricultural water management: present status and challenges. *Irrigation Science*, **26**(3), 223–237.
- Ha, W., Gowda, P. H., and Howell, T. A., 2013a. A review of down-scaling methods for remote sensing-based irrigation management: part I. *Irrigation Science*, **31**, 831–850.
- Ha, W., Gowda, P. H., and Howell, T. A., 2013b. A review of potential image fusion methods for remote sensing-based irrigation management: part II. *Irrigation Science*, **31**, 851–869.
- Hatfield, J. L., Quisenberry, J. E., and Dilbeck, R. E., 1987. Use of canopy temperatures to identify water conservation in cotton germplasm. *Crop Science*, **27**, 269–273.
- Heilman, J. L., Heilman, W. E., and Moore, D. G., 1982. Evaluating the crop coefficient using spectral reflectance. *Agronomy Journal*, **74**(6), 967–971.
- Herwitz, S. R., Johnson, L. F., Dunagan, S. E., Higgins, R. G., Sullivan, D. V., Zheng, J., Lobitz, B. M., Leung, L. G., Gallmeyer, B., Aoyagi, M., Slye, R. E., and Brass, J., 2004. Imaging from an unmanned aerial vehicle: agricultural surveillance and decision support. *Computers and Electronics in Agriculture*, **44**, 49–61.
- Howell, T. A., Meek, D. W., Phene, C. J., Davis, K. R., and McCormick, R. L., 1984. Automated weather data collection for research on irrigation scheduling. *Transactions of ASAE*, **27**(2), 386–391. 396.
- Howell, T. A., Schneider, A. D., Dusek, D. A., Marek, T. A., and Steiner, J. L., 1995. Calibration and scale performance of Bushland weighing lysimeters. *Transactions of the ASABE*, **38**(4), 1019–1024.
- Hunsaker, D. J., Barnes, E. M., Clarke, T. R., Fitzgerald, G. J., and Pinter, P. J., Jr., 2005. Cotton irrigation scheduling using remotely sensed and FAO-56 basal crop coefficients. *Transactions of ASAE*, **48**(4), 1395–1407.
- Hunsaker, D. J., Fitzgerald, G. J., French, A. N., Clarke, T. R., Ottman, M. J., and Pinter, P. J., Jr., 2007. Wheat irrigation management using multispectral crop coefficients: I crop evapotranspiration prediction. *Transactions of the ASABE*, **50**(6), 2017–2033.
- Idso, S. B., Jackson, R. D., and Reginato, R. J., 1977. Remote sensing of crop yields. *Science*, **196**, 19–25.

- Idso, S. B., Jackson, R. D., Pinter, P. J., Reginato, R. J., and Hatfield, J. L., 1981. Normalizing the stress degree day for environmental variability. *Agricultural Meteorology*, **24**, 45–55.
- Jackson, R. D., 1984. Remote sensing of vegetation characteristics for farm management. *Proceedings of SPIE*, **475**, 81–96.
- Jackson, T. J., and Hsu, A. Y., 2001. Soil moisture and TRMM microwave imager relationships in the Southern Great Plains 1999 (SGP99) experiment. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 1632–1642.
- Jackson, R. D., Idso, S. B., Reginato, R. J., and Pinter, P. J., Jr., 1981. Canopy temperatures as a crop water stress indicator. *Water Resources Research*, **17**(4), 1133–1138.
- Jackson, R. D., Kustas, W. P., and Choudhury, B. J., 1988. A reexamination of the crop water stress index. *Irrigation Science*, **9**, 309–317.
- Jackson, T. J., Le Vine, D. M., Swift, C. T., Schmugge, T. J., and Schiebe, F. R., 1995. Large area mapping of soil moisture using the ESTAR passive microwave radiometer in Washita'92. *Remote Sensing of Environment*, **53**, 27–37.
- Jackson, T. J., Le Vine, D. M., Hsu, A. Y., Oldak, A., Starks, P. J., Swift, C. T., Isham, J. D., and Haken, M., 1999. Soil moisture mapping at regional scales using microwave radiometry: the Southern Great Plains hydrology experiment. *IEEE Transactions on Geoscience and Remote Sensing*, **37**(5), 2136–2151.
- Jackson, T. J., Bindlish, R., Gasiewski, A. J., Stankov, B., Klein, M., Njoku, E. G., Bosch, D., Coleman, T. L., Laymon, C., and Starks, P., 2005. Polarimetric scanning radiometer C and X band microwave observations during SMEX03. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 2418–2430.
- Jackson, T. J., Moran, M. S., and O'Neill, P. E., 2008. Introduction to soil moisture experiments 2004 (SMEX04). *Remote Sensing of Environment*, **112**, 301–303.
- Jensen, M. E., 1968. Water consumption by agricultural plants. In Kozlowski, T. T. (ed.), *Water Deficits and Plant Growth*. New York: Academic, Vol. II, pp. 1–22.
- Jensen, M. E., Burman, R. D., and Allen, R. G. (eds.), 1990. *Evaporation and irrigation water requirements*. New York: American Society of Civil Engineers. ASCE Manuals and Reports on Eng. Practices No. 70, p. 360.
- Johnson, L., and Scholasch, T., 2005. Remote sensing of shaded areas in vineyards. *HortTechnology*, **15**(4), 859–863.
- Jones, H. G., 1992. *Plants and Microclimate*, 2nd edn. Cambridge, UK: Cambridge Press.
- Kim, Y., Evans, R. G., and Iversen, W. M., 2008. Remote sensing and control of an irrigation system using a wireless sensor network. *IEEE Transactions on Instrumentation and Measurement*, **57**(7), 1379–1387.
- Kim, Y., Evans, R. G., and Iversen, W. M., 2009. Evaluation of closed-loop site-specific irrigation with wireless sensor network. *Journal of Irrigation and Drainage Engineering*, **135**(1), 25–31.
- Kimball, B. A., LaMorte, R. L., Pinter, P. J., Jr., Wall, G. W., Hunsaker, D. J., Adamsen, F. J., Leavitt, S. W., Thompson, T. L., Matthias, A. D., and Brooks, T. J., 1999. Free-air CO₂ enrichment and soil nitrogen effects on energy balance and evapotranspiration of wheat. *Water Resources Research*, **35**, 1179–1190.
- King, B. A., Wall, R. W., and Wall, L. R., 2000. *Supervisory Control and Data Acquisition System for Closed-Loop Center Pivot Irrigation*. ASABE Paper No. 002020. St. Joseph: ASABE.
- Kustas, W. P., and Norman, J. M., 1999. Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. *Agricultural and Forest Meteorology*, **94**, 13–29.
- Kustas, W. P., Norman, J. M., Schmugge, T. J., and Anderson, M. C., 2004. Mapping surface energy fluxes with radiometric temperature. In Quattrochi, D., and Luvall, J. (eds.), *Thermal Remote Sensing in Land Surface Processes*. Boca Raton: CRC Press, pp. 205–253.
- Kustas, W. P., Agam, N., Anderson, M. C., Li, F., and Colaizzi, P. D., 2007. Potential errors in the application of thermal-based energy balance models with coarse resolution data. In Neale, C. M. U., Owe, M., and D'Urso, G. (eds.), *Remote Sensing for Agriculture, Ecosystems, and Hydrology IX*. Proc. of SPIE, Vol. 6742, doi: 10.1117/12.737776. 674208, (2007) 0277-786X/07/\$18.
- Mazahrir, N. T., Katbeh-Bader, N., Evett, S. R., Ayars, J. E., and Trout, T. J., 2008. Field calibration accuracy and utility of four down-hole water content sensors. *Vadose Zone Journal*, **7**, 992–1000.
- Miranda, F. R., Yoder, R., and Wilkerson, J. B. 2003. *A Site-Specific Irrigation Control System*. ASABE Paper No. 031129. St. Joseph: American Society of Agricultural and Biological Engineers.
- Möller, M., Alchanatis, V., Cohen, Y., Meron, M., Tsipris, J., Naor, A., Ostrovsky, V., Sprintsin, M., and Cohen, S., 2007. Use of thermal and visible imagery for estimating crop water status of irrigated grapevine. *Journal of Experimental Botany*, **58**(4), 827–838.
- Monteith, J. L., 1965. Evaporation and environment. In Fogg, G. E. (ed.), *Symposium of the Society for Experimental Biology, The State and Movement of Water in Living Organisms*. New York: Academic, Vol. 19, pp. 205–234.
- Moran, M. S., 1994. Irrigation management in Arizona using satellites and airplanes. *Irrigation Science*, **15**, 35–44.
- Moran, M. S., Clarke, T. R., Inoue, Y., and Vidal, A., 1994. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote Sensing of Environment*, **49**, 246–263.
- Moran, M. S., Mass, S. J., and Pinter, P. J., Jr., 1995. Combining remote sensing and modeling for estimating surface evaporation and biomass production. *Remote Sensing Reviews*, **12**(4), 335–353.
- Moran, M. S., Inoue, Y., and Barnes, E. M., 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment*, **61**, 319–346.
- Neale, C. M. U., Jayanthi, H., and Wright, J. L., 2003. Crop and irrigation water management using high-resolution airborne remote sensing. In *Proceedings of ICID Workshop Remote Sensing of ET for Large Regions*, CD-ROM. New Delhi: International Commission on Irrigation and Drainage.
- Norman, J. M., Kustas, W. P., and Humes, K. S., 1995. A two-source approach for estimating soil and vegetation energy fluxes from observations of directional radiometric surface temperature. *Agricultural and Forest Meteorology*, **77**, 263–293.
- Norman, J. M., Anderson, M. C., Kustas, W. P., French, A. N., Mecikalski, J., Torn, R., Diak, G. R., Schmugge, T. J., and Tanner, B. C. W., 2003. Remote sensing of surface energy fluxes at 101-m pixel resolutions. *Water Resources Research*, **39**(8), 1221.
- O'Shaughnessy, S. A. and Evett, S. R., 2007. IRT wireless interface for automatic irrigation scheduling of a center pivot system. In *Conference Proceedings of the 28th Annual International Irrigation Show*, San Diego Convention Center, San Diego, December 9–11, 2007. Reston: Irrigation Association, pp. 176–186.
- O'Shaughnessy, S. A., and Evett, S. R., 2008. Integration of wireless sensor networks into moving irrigation systems for automatic irrigation scheduling. In *2008 ASABE Annual International Meeting*, Rhode Island Convention Center, Providence, Rhode Island, June 29–July 2, 2008. American Society of Agricultural and Biological Engineers. Paper No. 083452.

- O'Shaughnessy, S. A., Evett, S. R., Colaizzi, P. D., and Howell, T. A., 2008. Soil water measurement and thermal indices for center pivot irrigation scheduling. In *Proceedings of the 2008 Irrigation Show and Conference*, November 2–4, 2008. Anaheim: Irrigation Association, Reston.
- Penman, H. L., 1948. Natural evaporation from open water, bare soil, and grass. *Proceedings of the Royal Society of London*, **A193**, 120–146.
- Peters, R. T., and Evett, S. R., 2004. Modeling diurnal canopy temperature dynamics using one time-of-day measurements and a reference temperature curve. *Agronomy Journal*, **96**, 1553–1561.
- Peters, R. T., and Evett, S. R., 2007. Spatial and temporal analysis of crop stress using multiple canopy temperature maps created with an array of center-pivot-mounted infrared thermometers. *Transactions of the ASABE*, **50**(3), 919–927.
- Peters, R. T., and Evett, S. R., 2008. Automation of a center pivot using the temperature-time-threshold method of irrigation scheduling. *Journal of Irrigation and Drainage Engineering*, **134**(3), 286–291.
- Phene, C. J., Howell, T. A., and Sikorski, M. D., 1985. A traveling trickle irrigation system. In Hillel, D. (ed.), *Advances in Irrigation*. Orlando, FL: Academic, Vol. 3, pp. 1–49.
- Plant, R. E., Munk, D. S., Roberts, B. R., Vargas, R. L., Rains, D. W., Travis, R. L., and Huttmacher, R. B., 2000. Relationships between remotely sensed reflectance data and cotton growth and yield. *Transactions of ASAE*, **43**(3), 535–546.
- Postel, S., 1999. *Pillar of Sand, Can the Irrigation Miracle Last?* Worldwatch Institute. New York: W.W. Norton, p. 312.
- Sadler, E. J., Bauer, P. J., Busscher, W. J., and Millen, J. A., 2000. Site-specific analysis of a droughted corn crop: II Water use and stress. *Agronomy Journal*, **92**(3), 403–410.
- Sadler, E. J., Camp, C. R., and Evans, R. G., 2007. New and future technology. In Lascano, R. J., and Sojka, R. E. (eds.), *Irrigation of Agricultural Crops*, 2nd edn. Madison: American Society of Agronomy/Crop Science Society of America/Soil Science Society of America. Agronomy Monograph, Vol. 30, pp. 609–626.
- Sela, E., Cohen, Y., Alchanatis, V., Saranga, Y., Cohen, S., Möller, M., Meron, M., Bosak, A., Tsipris, J., and Orollov, V., 2006. Use of thermal imaging for estimating and mapping crop water stress in cotton. In *6th European Conference on Precision Agriculture*, 3–6 June 2006, Skiathos.
- Shock, C. C., David, R. J., Shock, C. A., and Kimberling, C. A. (1999) Innovative, automatic, low-cost reading of Watermark soil moisture sensors. In *Proceedings of the 1999 Irrigation Association Technical Conference*. Falls Church: Irrigation Association, pp. 147–152.
- Todd, R. W., Evett, S. R., and Howell, T. A., 2000. The Bowen ratio-energy balance method for estimating latent heat flux of irrigated alfalfa evaluated in a semi-arid, advective environment. *Agricultural and Forest Meteorology*, **103**, 335–348.
- Twine, T. E., Kustas, W. P., Norman, J. M., Cook, D. R., Houser, P. R., Meyers, T. P., Prueger, J. H., Starks, P. J., and Wesley, M. L., 2000. Correcting eddy-covariance flux underestimates over a grassland. *Agricultural and Forest Meteorology*, **103**, 279–300.
- Upchurch, D. R., Wanjura, D. F., Burke, J. J., and Mahan, J. R., 1996. Biologically-identified optimal temperature interactive console (BIOTIC) for managing irrigation. United States Patent 5,539,637, July 23, 1996.
- Verhoest, N. E. C., Lievens, H., Wagner, W., Álvarez-Mozos, J., Moran, M. S., and Mattia, F., 2008. On the soil roughness parameterization problem in soil moisture retrieval of bare surfaces from synthetic aperture radar. *Sensors*, **8**, 4213–4248.
- Wall, R. W., and King, B. A., 2004. Incorporating plug and play technology into measurement and control systems for irrigation management. ASABE Paper No. 042189. St. Joseph: American Society of Agricultural and Biological Engineers.
- Xiang, H., and Tian, L., 2007. Autonomous aerial image georeferencing for an UAV-Based data collection platform using integrated navigation system. Paper Number: 073046. St. Joseph: American Society of Agricultural and Biological Engineers.
- Zhan, X., Crow, W. T., Jackson, T. J., and O'Neill, P., 2008. Improving space-borne radiometer soil moisture retrievals with alternative aggregation rules for ancillary parameters in highly heterogeneous vegetated areas. *IEEE Transaction on Geoscience and Remote Sensing*, **2**, 261–265.

Cross-references

[Calibration, Optical/Infrared Passive Sensors](#)
[Commercial Remote Sensing](#)
[Crop Stress](#)
[Emerging Technologies, Sensor Web](#)
[Microwave Radiometers](#)
[Observational Platforms, Aircraft, and UAVs](#)
[Observational Systems, Satellite](#)
[Optical/Infrared, Radiative Transfer](#)
[Precision Agriculture](#)
[Rainfall](#)
[Remote Sensing, Physics and Techniques](#)
[Soil Moisture](#)
[Soil Properties](#)
[Surface Radiative Fluxes](#)
[Surface Water](#)
[Thermal Radiation Sensors \(Emitted\)](#)
[Vegetation Indices](#)
[Water and Energy Cycles](#)
[Water Resources](#)

L

LAND SURFACE EMISSIVITY

Alan Gillespie
Department of Earth and Space Sciences, University
of Washington, Seattle, WA, USA

Definitions

Land surface emissivity (LSE). Average emissivity of an element of the surface of the Earth calculated from measured radiance and land surface temperature (LST) (for a complete definition, see Norman and Becker, 1995).

Atmospheric window. A spectral wavelength region in which the atmosphere is nearly transparent, separated by wavelengths at which atmospheric gases absorb radiation. The three pertinent regions are “visible/near-infrared” ($\sim 0.4\text{--}2.5\ \mu\text{m}$), mid-wave infrared ($\sim 3\text{--}5\ \mu\text{m}$) and long-wave infrared ($\sim 8\text{--}14\ \mu\text{m}$).

Blackbody. An ideal material absorbing all incident energy or emitting all thermal energy possible. A cavity with a pinhole aperture approximates a blackbody.

Brightness temperature. The temperature of a blackbody that would give the radiance measured for a surface.

Color temperature. Temperature satisfying Planck’s law for spectral radiances measured at two different wavelengths.

Contrast stretch. Mathematical transform that adjusts the way in which acquired radiance data translate to the black/white dynamic range of the display monitor.

Emissivity ε . The efficiency with which a surface radiates its thermal energy.

Irradiance. The power incident on a unit area, integrated over all directions (W m^{-2}).

Graybody. A material having constant but non-unity emissivity.

Long-wave infrared (LWIR). For most terrestrial surfaces ($\sim 340\ \text{K}$ to $\sim 240\ \text{K}$), peak thermal emittance occurs in the LWIR ($\sim 8\text{--}14\ \mu\text{m}$).

Mid-infrared (MIR). Forest fires ($\sim 1,000\text{--}600\ \text{K}$) have peak thermal emittances in the MIR ($\sim 3\text{--}5\ \mu\text{m}$).

Noise equivalent Δ temperature ($NE\Delta T$). Random measurement error in radiance propagated through Planck’s law to give the equivalent uncertainty in temperature.

Path radiance $S_{\uparrow}$. The power per unit area incident on a detector and emitted upward from within the atmosphere ($\text{W m}^{-2}\ \text{sr}^{-1}$).

Planck’s law. A mathematical expression relating spectral radiance emitted from an ideal surface to its temperature (Equation 1, in the entry *Land Surface Temperature*).

Radiance. The power per unit area from a surface directed toward a sensor, in units of $\text{W m}^{-2}\ \text{sr}^{-1}$.

Reflectivity ρ . The efficiency with which a surface reflects energy incident on it.

Reststrahlen bands. Spectral bands in which there is a broad minimum of emissivity associated in silica minerals with interatomic stretching vibrations of Si and O bound in the crystal lattice.

SEBASS. Spatially Enhanced Broadband Array Spectrograph System, a hyperspectral TIR imager (Hackwell et al., 1996).

Short-wave infrared (SWIR). Erupting basaltic lavas ($\sim 1,400\ \text{K}$) have their maximum thermal emittance at $\sim 2.1\ \mu\text{m}$ in an atmospheric window at $0.4\text{--}2.5\ \mu\text{m}$. Part of this spectral region ($1.4\text{--}2.5\ \mu\text{m}$) is called the SWIR.

Sky irradiance $I_{\downarrow}$. The irradiance on the Earth’s surface originating as thermal energy radiated downward by the atmosphere (W m^{-2}) (spectral irradiance: $\text{W m}^{-2}\ \mu\text{m}^{-1}$).

Spectral radiance L . Radiance per wavelength, in units of $\text{W m}^{-2}\ \mu\text{m}^{-1}\ \text{sr}^{-1}$.

Thermal infrared (TIR). Thermal energy is radiated from a body at frequencies or wavelengths in proportion to its temperature. The wavelengths for which this radiant energy

is significant for most terrestrial surfaces ($\sim 1.4\text{--}14\ \mu\text{m}$) are longer than the wavelength of visible red light and hence are known as thermal infrared. The TIR is subdivided into three ranges (LWIR, MIR, SWIR) for which the atmosphere is transparent (atmospheric “windows”) so that the energy can be measured from space.

Introduction

Thermal emissivity ε is the efficiency with which a surface emits its stored heat as thermal infrared (TIR) radiation. It is useful to know because it indicates the composition of the radiating surface and because it is necessary as a control in atmospheric and energy-balance models, since it must be known along with brightness temperature to establish the heat content of the surface. The first practical demonstration of multispectral TIR imaging for compositional mapping was from a NASA airborne scanner flown over Utah (Kahle and Rowan, 1980).

Emissivity differs from wavelength to wavelength, just as reflectivity ρ does in the spectral region of reflected sunlight ($0.4\text{--}2.5\ \mu\text{m}$). Emissivity is defined as

$$\varepsilon(\lambda) = \frac{L(\lambda, T)}{B(\lambda, T)} \quad (1)$$

where L is the measured spectral radiance and B is the theoretical blackbody spectral radiance for a surface with a skin temperature T . B is given by Planck’s law which, together with the basic physics of TIR radiative transfer, is discussed in the entry *Land Surface Temperature* (LST).

Unlike T , which is a variable property of a surface controlled by the heating history and not directly by composition, $\varepsilon(\lambda)$ is independent of T and is a function directly of composition. Furthermore, $\varepsilon(\lambda)$ in the TIR wavelengths ($3\text{--}14\ \mu\text{m}$) responds to different aspects of composition than reflectivity $\rho(\lambda)$ at $0.4\text{--}2.5\ \mu\text{m}$. In general, ρ at wavelengths $0.4\text{--}2.5\ \mu\text{m}$ is controlled by the amounts of iron oxides, chlorophyll, and water on the surface; ε in the TIR is controlled more by the bond length of Si and O in silicate minerals. Examples of emissivity spectra are given in Figure 1.

TIR spectroscopy is especially important because silicate minerals are the building blocks of the geologic surface of Earth, and their presence and amounts can be inferred only indirectly at shorter wavelengths. Thus TIR spectroscopy is complementary to spectroscopy of reflected sunlight. Good summaries of TIR spectroscopy and its significance in terms of surface composition may be found in Lyon (1965), Hunt (1980), and Salisbury and D’Aria (1992). A good introduction to spectral analysis may be found in Clark et al. (2003).

Figure 1 shows daytime and nighttime false-color composite images of spectral radiance from a sparsely vegetated part of Death Valley, California, enhanced using a decorrelation contrast stretch (Soha and Schwartz, 1978; Gillespie et al., 1986). This stretch emphasizes the emissivity component of the signal, shown as color, and

de-emphasizes the temperature, shown as dark/light intensity. In addition to composition, the daytime image gives a good sense of topography, because sunlit slopes are warmer than shadowed slopes. In the nighttime image, most temperature effects are subdued, and the image closely resembles the Land Surface Emissivity (LSE) alone.

Exceptions include standing water, which is cooler than the land during the day but warmer at night. Standing water (C) in the floor of Death Valley shows dark green in the daytime image but light pink in the nighttime image. Vegetation (A) appears dark in the daytime image, when it is cooling its canopy by evapotranspiration. The toe of an alluvial fan (B) appears darker at night, when soil moisture rises to the surface and evaporates.

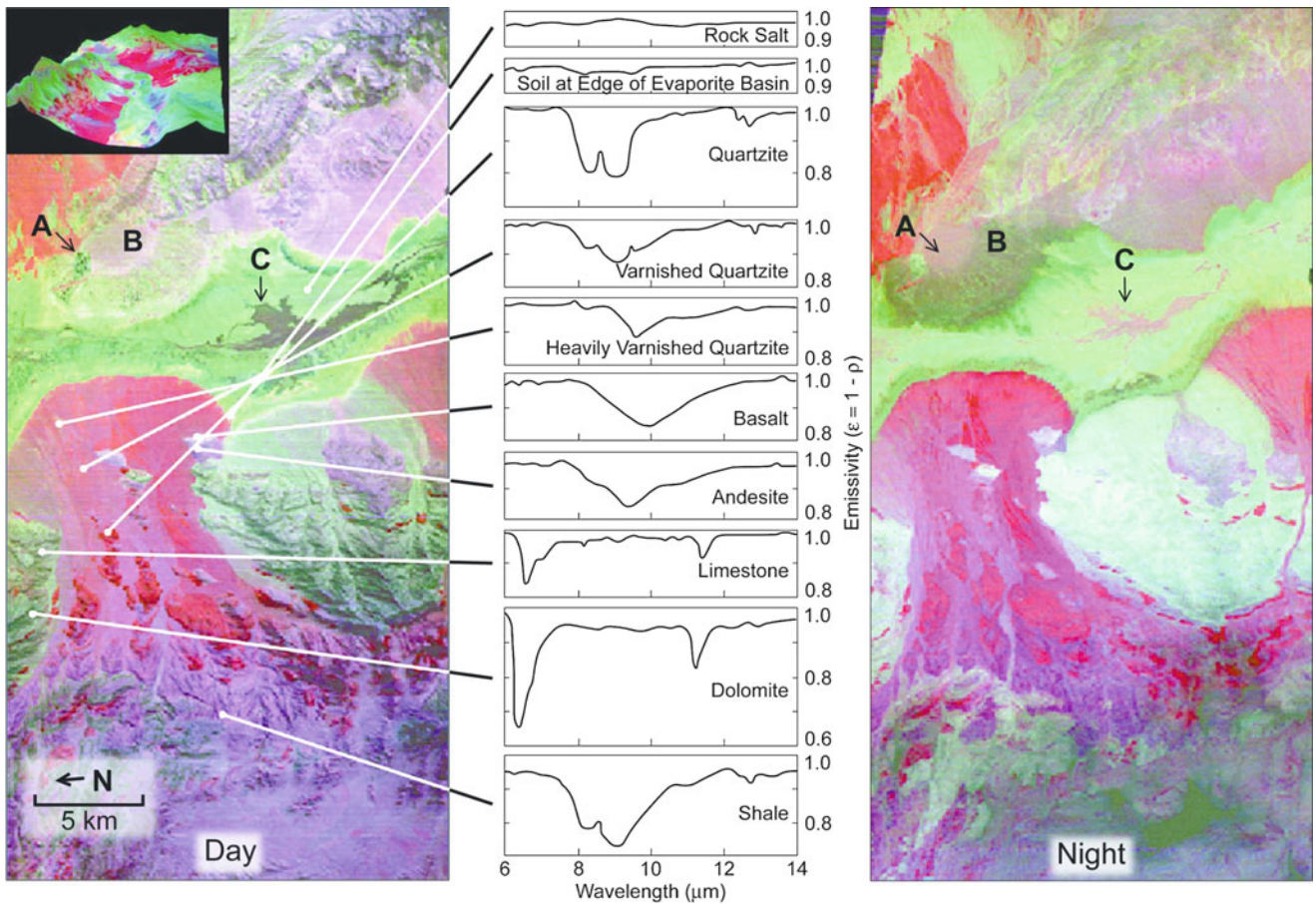
The colors in Figure 1 indicate rock type. For example, the emissivity of quartzite is low (~ 0.8) at 8.3 and $9.1\ \mu\text{m}$ (blue and green) but high at $10.4\ \mu\text{m}$ (red); therefore, it is displayed as red. Other rock types and display colors can be understood by comparing the images and emissivity spectra in Figure 1.

The discussion below focuses on algorithms designed to recover emissivities from remotely sensed spectral radiance data. Figure 2, of a desert landscape, compares spectral radiance to temperature and emissivity images recovered from it. Also shown are emissivity spectra of vegetation and the geologic substrate. As explained in the entry *Land Surface Temperature*, temperature and emissivity recovery is an underdetermined problem, and dozens of approaches have been proposed and published that break down the indeterminacy. These fall in four classes: deterministic algorithms that solve for *LST* and *LSE* exactly, algorithms that recover the shape of the LSE spectrum only, model approaches that make key assumptions, and algorithms that attempt also to scale or calibrate the normalized spectra to their actual emissivity values.

In evaluating the algorithms, it is useful to ask how accurately it is necessary to recover LSE and LST. For example, many analytic algorithms that seek to identify surface composition rely not so much on actual emissivity values, but on the central wavelengths of emissivity minima (e.g., reststrahlen bands), which can be diagnostic for many rocks and minerals. If this is your goal, it may not be necessary to scale the spectra, relying instead on the simpler algorithms that just recover spectral shape.

Errors in LST may affect some algorithms by warping the spectra over several μm of wavelength. This happens because the shape of the Planck function changes with temperature (*Land Surface Temperature*, Figure 2). A 5 K error at 300 K, for example, will cause a slope in the recovered emissivity spectrum of 0.05 from 8 to $14\ \mu\text{m}$. However, the sharp mineralogical features ($\sim 0.2\text{--}0.5\ \mu\text{m}$ wide) are readily distinguished against this distorted continuum.

The TIR is commonly a difficult spectral region in which to measure spectral radiance, and the images are typified by a low signal–noise ratio. This ratio is commonly represented by the “noise equivalent Δ temperature” or



Land Surface Emissivity, Figure 1 Airborne thermal infrared multispectral scanner (TIMS: Palluconi and Meeks, 1985) false-color TIR radiance images of Death Valley, California (RGB = 10.4, 9.1, 8.3 μm). Letters A, B, and C indicate sites discussed in the text. *Central column* shows laboratory spectra for field samples. *Inset* shows similar ASTER image “draped” over topography, looking north up Death Valley. The TIMS images cross the central part of the ASTER footprint (Courtesy Harold Lang and Anne Kahle, JPL).

NEAT, which is the temperature difference corresponding to the standard deviation of the radiance within a homogeneous, isothermal scene region. For TIR imagers such as ASTER, $NEAT_{300K} \approx 0.25$ K. Also for ASTER, the NEAT, atmospheric correction, and radiometric calibration all introduce errors of about the same size, leading to a total uncertainty in the recovered LST of about 1.5 K and in the LSE of ~ 0.015 .

Deterministic solutions for emissivity

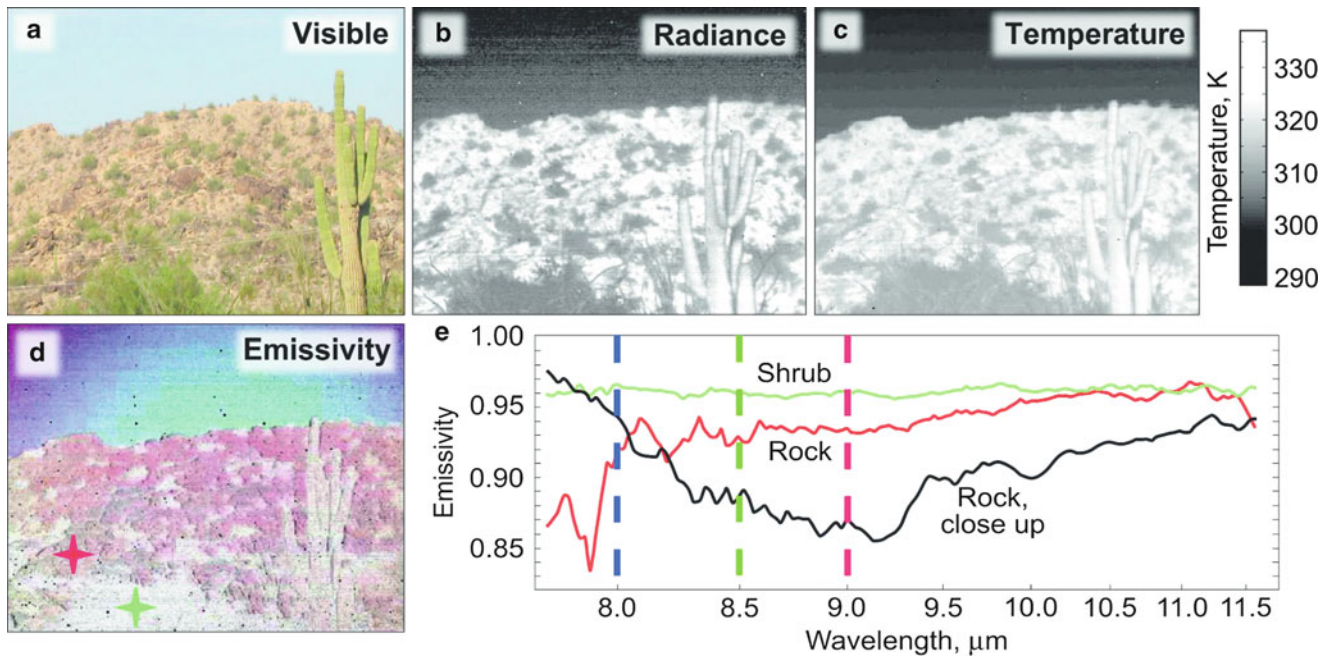
As discussed in *Land Surface Temperature*, recovering both LST and LSE from a single image is underdetermined. In principle, this problem can be removed by increasing the number of images acquired for the same scene. For each n -channel image, after atmospheric compensation, there are $n + 1$ unknowns, but only n measurements; for two images of the same scene, there are $n + 2$ unknowns, but $2n$ measurements (assuming LST has changed but LSE has remained constant). Therefore, a two-channel image taken at two different times is

deterministic. It is additionally necessary that the LST be significantly different between acquisitions.

Two-time, two-channel approach

If well-registered multispectral day–night radiance measurements are available, it is possible to determine T and ϵ uniquely (Watson, 1992a). Although this approach is esthetic, for most TIR data, the recovered temperatures and emissivities tend to be imprecise. For example, for image channels at 8 and 12 μm , day–night temperatures of 290 and 310 K, and for $NEAT = 0.3$ K, recovered LST would have an uncertainty of ≥ 20 K. This arises because of the flat shape of the Planck curve in the spectral range around 300 K.

Wan (1999) showed that using an image channel in the 3–5 μm window, where the slope of the Planck function is steep, can improve the precision greatly and used the day–night algorithm to make a standard MODIS LST product. However, for daytime data, reflected sunlight at 3–5 μm must be accounted for (see *Land Surface Temperature*, Figure 3). Furthermore, acquiring data 12 h or more apart



Land Surface Emissivity, Figure 2 TIR images and spectra, South Mountain, Arizona, looking SE. (a) Natural color; (b) TIR radiance at 9 μm; (c) brightness temperature; (d) emissivity (RGB = 8, 8.5 and 9 μm, respectively); (e) emissivity spectra measured with the TELOPS, Inc., FIRST hyperspectral imaging spectrometer, August 8, 2007. The shrub spectrum was taken from the site in d marked the green cross; the rock spectrum from the red cross. Differences in the “rock” spectra likely relate to differences in the pixel field of view and exact location, and in the length of the atmospheric path between the sample and sensor.

adds complexity because the scene may have changed between images, for example, because of dew.

It is also advantageous to use more than two channels, in which case the inversion for LST and LSE is overdetermined. This has the advantage of reducing the impact of measurement errors. The exaggeration of measurement error in this otherwise esthetic technique will become less severe as high-precision imagers such as SEBASS ($NEAT_{300K, 11\mu m} < 0.05$ K; Hackwell et al., 1996) become widely available.

Emissivity bounds method

Jaggi et al. (1992) observed that for every pixel and every channel i there exists a locus of (T, ϵ_i) vectors that are possible solutions for the modified Planck equation (Equation 1, *Land Surface Temperature*). Because T must be the same for all image channels, some (T, ϵ_i) pairs can be ruled out as candidate solutions. The range of solutions is even more limited if ϵ and/or T can be restricted a priori. For the land surface, it is commonly possible to assume that $0.8 < \epsilon < 1.0$, for example.

This elegant approach is not truly deterministic, because it requires assumed limits to ϵ and/or T . However, it requires no empirical assumptions. The technique does not appear to have been widely used, perhaps because it does not identify the most probable values of ϵ or T , only possible ranges. In practice, performance depends on

how well emissivity limits are known a priori, and implementation would probably require some sort of image classification to establish them closely.

Spectral-shape solutions

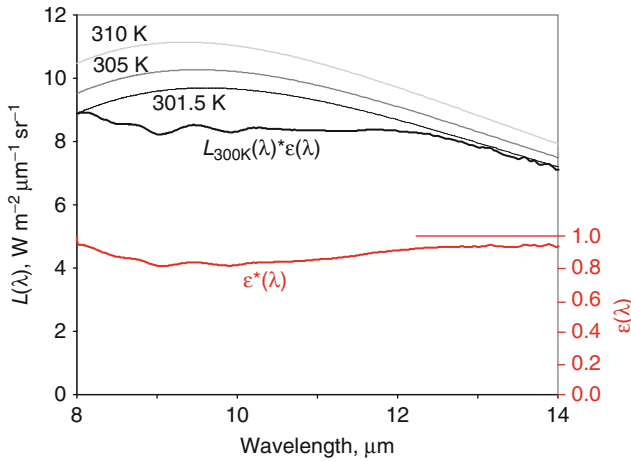
Although it is not possible to invert the modified Planck equation for both ϵ and T without external constraints, it is possible to estimate spectral shape for ϵ , at the expense of T and of the amplitude of the recovered spectrum, that is, the recovered spectra are essentially normalized, so that only relative amplitudes (wavelength to wavelength) are known. This is nevertheless useful, since composition is generally determined from spectral shape, and not the absolute amplitudes.

Ratio methods

Watson (1992b) observed that ratios of spectrally adjacent channels i and j described spectral shape accurately, provided that T could be estimated even roughly:

$$\frac{\epsilon_j}{\epsilon_i} = \frac{L_j \lambda_i^5 (\exp(c_2/(\lambda_i T)) - 1)}{L_i \lambda_j^5 (\exp(c_2/(\lambda_j T)) - 1)} \quad (2)$$

(c_2 is a constant from Planck's law, Equation 1, *Land Surface Temperature*). To calculate the ϵ ratios, it is necessary first to approximate the temperature T from the measured radiances L_i and L_j . If ϵ can be estimated



Land Surface Emissivity, Figure 3 Emissivity $\varepsilon(\lambda)$ and spectral radiance spectra $L(\lambda)$ for basalt at 300 K. $L(\lambda)$ was calculated as the product of measured $\varepsilon(\lambda)$ and a 300 K blackbody ($B(\lambda)$) spectrum. In “Planck draping,” blackbody spectra are calculated for successively lower temperatures (e.g., 310, 305 and 301.5 K, above) until $\varepsilon_{\max} B(\lambda) = L(\lambda)$ at some wavelength. The maximum emissivity, $\varepsilon_{\max}$ must be estimated, usually as a value near 0.95 as in the example shown. $\varepsilon^*(\lambda)$, the recovered $\varepsilon(\lambda)$, is calculated as $L(\lambda)/B(\lambda)$. Both it and the found LST (301.5 K, above) will be inaccurate unless the Planck functions are scaled correctly by $\varepsilon_{\max}$. In the example shown, LST is in error by 1.5 K. The error warps $\varepsilon^*(\lambda)$ slightly.

within ± 0.075 , the uncertainty in T is ± 5 K, and the ε ratios can be estimated with an average error of ~ 0.007 (this estimate does not include the effects of measurement error).

Becker and Li (1990) proposed a similar approach they called the “temperature-independent spectral indices” (TISI) method. TISI begins with the observation (Slater, 1980) that Planck’s law may be represented by

$$B_k(T_s) = \alpha_k(T_o) T^{n_k(T_o)} \quad (3)$$

where B is the spectral radiance in image channel k for a blackbody at temperature T_s and T_o is a reference temperature. Constants n_k and α_k are given by

$$n_k(T_o) = \frac{c_2}{\lambda_k T_o} \left(1 + \frac{1}{\exp(c_2/\lambda_k T_o) - 1} \right); \quad (4)$$

$$\alpha_k(T_o) = \frac{B_k(T_o)}{T_o^{n_k(T_o)}}$$

(Dash, 2005). The land-leaving spectral radiance L_k , corrected for atmospheric absorption and path radiance but not down-welling spectral irradiance $L_k^{\downarrow}$, is thus

$$L_k = \varepsilon_k \alpha_k T_s^{n_k} C_k; \quad C_k = 1 + \frac{(1 - \varepsilon_k) L_k^{\downarrow}}{\varepsilon_k B_k(T_s)} \quad (5)$$

where C_k is spatially variable and atmosphere specific. The TISI is found by rationing spectral radiances for image channels i and j :

$$\frac{L_i^{a_i}}{L_j^{a_j}} = \frac{\varepsilon_i^{a_i} \alpha_i^{a_i} T_s^{n_i a_i} C_i^{a_i}}{\varepsilon_j^{a_j} \alpha_j^{a_j} T_s^{n_j a_j} C_j^{a_j}} \quad (6)$$

Here a_i is defined as n_i^{-1} (and $a_j = n_j^{-1}$), chosen to make Equation 6 independent of T . Since for a wide range of temperatures the C ratio is close to unity, TISI is then

$$TISI_{i,j} = \left[\frac{L_i}{\alpha_i} \right]^{1/n_i} \left[\frac{L_j}{\alpha_j} \right]^{-1/n_j} = \frac{\varepsilon_i^{1/n_i} C_i^{1/n_i}}{\varepsilon_j^{1/n_j} C_j^{1/n_j}} \approx \frac{\varepsilon_i^{1/n_i}}{\varepsilon_j^{1/n_j}} \quad (7)$$

The ratio spectra are insensitive to temperature, for normal terrestrial ranges. The approaches are adaptable for most sensors.

Alpha-residual method

The alpha-residual algorithm produces a relative emissivity spectrum that preserves spectral shape but, like the ratio methods, does not yield actual ε or T values. The alpha residuals are calculated utilizing Wien’s approximation of Planck’s law, which neglects the “ -1 ” term in the denominator. This makes it possible to linearize the approximation with logarithms, thereby separating λ and T :

$$\frac{c_2}{T} \approx \lambda_j \ln(\varepsilon_j) - \lambda_j \ln(L_j) + \lambda_j \ln(c_1) - 5\lambda_j \ln(\lambda_j) - \lambda_j \ln(\pi). \quad (8)$$

Here c_1 and c_2 are the constants defined in Planck’s law (Equation 1, *Land Surface Temperature*) and j is the image channel. Wien’s approximation introduces a systematic error in ε_j of $\sim 1\%$ at 300 K and 10 μm wavelength.

The next step is to calculate the means for the parameters of the linearized equation, summing over the n image channels:

$$\frac{c_2}{T} \approx \frac{1}{n} \sum_{j=1}^n \lambda_j \ln(\varepsilon_j) - \frac{5}{n} \sum_{j=1}^n \lambda_j \ln(\lambda_j) - \frac{1}{n} \sum_{j=1}^n \lambda_j \ln(L_j) + (\ln(c_1) - \ln(\pi)) \frac{1}{n} \sum_{j=1}^n \lambda_j. \quad (9)$$

The residual is calculated by subtracting the mean from the individual channel values. Collecting terms, a set of n equations is generated relating ε_i to L_i , independent of T :

$$\lambda_j \ln(\varepsilon_j) - \mu_\alpha = \lambda_j \ln(\lambda_j) - \frac{1}{n} \sum_{j=1}^n \lambda_j \ln(L_j) + \kappa_i. \quad (10a)$$

$$\kappa_i \approx 5\lambda_i \ln(\lambda_i) - \sum_{j=1}^n \lambda_j \ln(\lambda_j) - (\ln(c_1) - \ln(\pi))(\lambda_j - \bar{\lambda}) \quad (10b)$$

$$\mu_\alpha = \frac{1}{n} \sum_{j=1}^n \lambda_j \ln(\varepsilon_j); \quad \bar{\lambda} = \frac{1}{n} \sum_{j=1}^n \lambda_j \quad (10c)$$

Note that κ_i contains only terms which do not include the measured spectral radiances, L_i , and hence may be calculated from the constants. Although dependency on T has been eliminated, it has been replaced by the unknown μ_α , related to the mean emissivity, such that the total number of unknowns is unchanged. The components of the alpha-residual spectrum vary only with the measured radiances. They are defined as

$$\alpha_i \equiv \lambda_i \ln(\varepsilon_i) - \mu_\alpha \quad (11)$$

and are equivalent to the right-hand side of Equation 10a (α is defined differently than in the TISI method).

Model approaches

In this section, three algorithms distinguished by their model assumptions are described. The most specific requires that both a value of ε and the wavelength at which it occurs be known. The next requires only that the value be known. The third does not require the value of the emissivity to be known, only that the emissivity at two known wavelengths be the same.

The model emissivity (or reference channel) method (Kahle et al., 1980) assumes that the value of ε for one of the image channel's ref is constant and known a priori, reducing the number of unknowns to the number of measurements. First, the temperature is estimated using

$$T = \frac{c_2}{\lambda_{ref}} \left(\ln \left(\frac{c_1 \varepsilon_{ref}}{\pi L_{ref} \lambda_{ref}^5} \right) + 1 \right)^{-1} \quad (12)$$

Lyon (1965) suggested that, for most rocks, the maximum emissivity (ε_{max}) was commonly ~ 0.95 and occurred at the long-wavelength end of the 8–14 μm TIR window. This observation has been used to justify the assumption $\varepsilon_{ref} = \varepsilon_{max}$, typically for $10 < \lambda_{ref} < 12 \mu\text{m}$.

Blackbody spectral radiances B_j for the remainder of the channels are next calculated from T and Planck's law. The model emissivities are $\varepsilon_j = L_j/B_j$.

No single value of ε_{ref} is appropriate for all surfaces. For example, for vegetation, $\varepsilon_{max} \approx 0.983$; if the value is assumed to be 0.95, the emissivities will be underestimated, the spectrum warped, and T overestimated by ~ 2.3 K. Vegetation, snow, and water are all subject to this kind of error. Also, reststrahlen bands for some types of rocks, for example, peridotite, occur near 10 μm , and ε_{max} occurs at shorter wavelengths. For these rock types, the errors may be even greater. Nevertheless, the model emissivity approach is robust and has the virtue of simplicity. It produces reliable results for a wide range of surface materials.

Retaining the assumption $\varepsilon_{ref} = \varepsilon_{max}$ but allowing the reference channel to vary, pixel by pixel, allows the model emissivity approach to be accurate for a wider range of materials. This approach is called the normalized emissivity method (NEM) (Gillespie, 1985; Realmuto, 1990). First, the brightness temperature T_b is found for each image channel, using Planck's law. T_b differs from channel to channel only if ε_j does also, since the actual skin temperature must be the same. The channel j with the maximum T_b is also the channel for which the maximum ε_j occurs and becomes the reference channel. For 81 spectra evaluated by Hook et al. (1992), 58 % of the temperatures found by the NEM algorithm were accurate to within 1 K, compared to only 21 % of temperatures recovered using the model emissivity method.

Finding the maximum T_b has been called the "Planck draping" method (Figure 3). This approach has been used to estimate $\varepsilon(\lambda)$ from high-resolution radiance spectra collected by hyperspectral imagers such as SEBASS or by field spectrometers.

Instead of examining the same scene element at two different times and temperatures, as in the day–night method, the scene element may be measured at different wavelengths λ_i and λ_j , chosen such that $\varepsilon_i = \varepsilon_j$. In such a case, it is necessary to find T (the "color temperature," T_c ; see Equations 10 and 11, *Land Surface Temperature*) and only a single ε for the two channels, and the situation is deterministic (two measurements, L_i and L_j , and the two unknowns, T_c and $\varepsilon_i = \varepsilon_j$). As for the reference channel method, T_c can then be used to calculate a blackbody spectrum B , from which $\varepsilon(\lambda)$ can be found. This treatment has been called the "graybody emissivity method" (Barducci and Pippi, 1996).

The strength of the technique lies in its ability to recover emissivities even if the value of ε is unknown. The main weaknesses are that for imagers with only a few TIR channels, the basic requirement, $\varepsilon_i = \varepsilon_j$, is not met for much of the land surface, and it is not always possible to know λ_i and λ_j . If the assumption is valid, the accuracy for T is comparable to NEM, provided λ_i and λ_j are widely separated (e.g., Mushkin et al., 2005), but for most rock spectra, errors are ≥ 5 K. Barducci and Pippi (1996) proposed the graybody emissivity method for hyperspectral scanners, for which the basic requirement is more likely to be met.

Scaling approaches

Once relative spectra have been calculated, they can be calibrated to "absolute" emissivity provided a scaling factor is known. Applied to the ratio approach of Watson (1992b) or the TISI approach of Becker and Li (1990), this is basically the same as one of the model algorithms. However, scaling can also be done from empirical regression relating the shape of the emissivity spectrum to an absolute value at one wavelength. The regression is typically based on laboratory spectra of common scene components. More complex approaches also are possible: the first example given below combines the "two-channel,

two-time,” and TISI approaches to convert the relative TISI spectra to emissivities.

The hybrid TISI approach requires first that daytime and nighttime MIR and LWIR images be acquired and co-registered and that their TISI ratios be calculated. Essentially, there are four measurements ($L_{MIR,day}$, $L_{LWIR,day}$, $L_{MIR,night}$, and $L_{LWIR,night}$), four unknowns (ε_{MIR} , ε_{LWIR}), and one model assumption (the solar irradiance on the target). The MIR reflectivity is the complement of ε_{MIR} by Kirchhoff's law (for the complete mathematical development, see Dash, 2005). Using widely separated image channels improves the precision of T and ε recovery (e.g., Mushkin et al., 2005).

Alpha-derived emissivity (ADE) method

The ADE method (Kealy and Gabell, 1990; Hook et al., 1992; Kealy and Hook, 1993) is based on the alpha-residual approach. To recover ε_i , μ_α may be estimated via an empirical regression to the variance parameter v_α found for laboratory spectra:

$$v_\alpha = \frac{1}{n-1} \sum_{j=1}^n \alpha_j^2 \quad (13)$$

where α is defined in Equation 11. The best-fitting curves relating μ_α and v_α are of the form $\mu_\alpha = cv_\alpha^{1/x}$, where c and x are empirically determined coefficients ($c = -0.085$, $v_\alpha = 0.40$, and $r^2 = 0.935$ for ASTER).

Once the emissivities have been estimated, the temperature may be calculated using Planck's law. For 95 % of the library spectra, T was recovered within 1.6 K of the correct value, and Hook et al. (1992) showed that 67 % were accurate to within 1 K, compared to 58 % for the NEM.

The key innovation of the ADE approach is to utilize the empirical relationship between the average ε and a measure of the spectral contrast or complexity in order to restore the amplitude to the alpha-residual spectrum. The regression is based on the observation that, for a blackbody, the mean emissivity is unity and the spectral variance is zero. For minerals with reststrahlen bands or other emissivity features, the variance is greater than zero and, of course, the mean is less than unity. In use, the mean is predicted from the variance, which is calculated from the measured radiances.

Temperature–emissivity separation algorithm (TES)

The TES algorithm (Gillespie et al., 1998) uses a variant of the “minimum–maximum difference” or MMD approach of Matsunaga (1994) to scale relative emissivity spectra. TES is used to generate standard T and ε products from ASTER, but it has been generalized for different scanners. TES can work with as few as three channels provided the channel wavelengths are well chosen to capture the range of emissivities in scene spectra.

The MMD algorithm is related to the ADE algorithm, but is simpler. Whereas ADE utilizes the empirical

relationship between the mean emissivity $\bar{\varepsilon}$ and the variance of alpha-residual emissivities, MMD utilizes an assumed linear relationship between $\bar{\varepsilon}$ and the range of the emissivities themselves, represented by the maximum–minimum difference or *MMD*.

The MMD algorithm requires that the ε spectrum be estimated (e.g., using NEM) in order to calculate *MMD*, from which $\bar{\varepsilon}$ is predicted. The apparent spectrum is then rescaled according to this average, T is calculated, and the process is iterated until the change in T is less than the *NEAT*.

TES uses land-leaving spectral radiance and downwelling sky irradiance as input and provides a first guess for T and ε_j using the NEM algorithm. The correction for reflected sky irradiance is

$$L'_j = L_j - \frac{(1 - \varepsilon_j)}{\pi} I_\downarrow \quad (14)$$

where L_j is the ground-leaving spectral radiance, compensated for atmospheric absorption and path radiance, $I_\downarrow$ is the downwelling sky irradiance, and $(1 - \varepsilon_j)$ is the scene reflectivity (Kirchhoff's law). The NEM emissivities are then recalculated from L'_j and normalized:

$$\beta_j = \frac{\varepsilon_j}{\bar{\varepsilon}} \quad (15)$$

MMD is calculated from the β spectrum and used to predict ε_{min} (instead of $\bar{\varepsilon}$, as in the MMD approach), which is used for scaling:

$$\varepsilon_{min} = 0.994 - 0.687 MMD^{0.737}; \quad \varepsilon_j = \beta_j \left(\frac{\varepsilon_{min}}{\beta_{min}} \right) \quad (16)$$

After early 2009, a linear regression ($\varepsilon_{min} = 0.8625 MMD + 0.955$) was used for scaling in TES (Gustafson et al., 2006) in order to improve TES precision for low-contrast spectra in standard ASTER data products. The TES algorithm differs from the MMD approach in using a better estimate of the emissivity and in basing the “absolute” measure of emissivity on ε_{min} rather than $\bar{\varepsilon}$, a difference that results in less scatter of the data about the regressed line and, hence, improved performance (± 1.5 K; $\pm 0.015 \varepsilon$).

Classification-based algorithms

Classification approaches exploit the relationship between composition and ε and/or ρ to estimate ε pixel by pixel in at least one-image channel, generally in order to find T . T can then be used to calculate $\varepsilon(\lambda)$ in the other channels. Approaches that use channels in reflected sunlight (0.4–2.5 μm) require imagers with multiple, co-registered telescopes. They also make the assumption that TIR emissivities and visible–SWIR reflectivities are correlated. In some cases, for example, vegetation or water, the TIR emissivities can be predicted accurately; in others, for

example, many rocks, this assumption is less robust. Nevertheless, simply being able to distinguish rock and/or soil from vegetation can improve accuracy by 1–2 K. As an example, the NDVI approach (see Equation 9, *Land Surface Temperature*) makes use of co-registered visible red ($\sim 0.65 \mu\text{m}$) and near-infrared (NIR: $0.7\text{--}1.2 \mu\text{m}$) daytime image channels in order to recognize pixels that have a significant fraction of vegetation.

Conclusions

Only a fraction of published temperature–emissivity separation algorithms have been discussed here. (For an alternative summary, see Dash (2005).) Increasingly sophisticated approaches are being devised to improve on old treatments, for example, by using neural net technology to tune algorithms (e.g., Mao et al., 2008; Liang, 1997). However, the basic categories discussed above still apply.

For the most part, calibration inaccuracies, measurement uncertainty, and inaccurate atmospheric characterization all contribute to errors in the recovered LST and LSE. These errors are commonly as large as or larger than those attributable to the algorithms themselves, at least for the high-resolution imagers commonly used for Earth-surface studies. Therefore, algorithms themselves are now not the dominant factor limiting recovery accuracy. However, the next few years may see the introduction of a new generation of sensors, such as SEBASS, with dramatically improved measurement characteristics. In this case, atmospheric compensation may become the biggest source of uncertainty and deserving of attention. Likewise, the performance of some algorithms like the “two-time, two-channel” algorithm that now are strongly limited by measurement precision may improve relative to those algorithms that are limited by different factors, such as TES with its empirical regression of $\varepsilon_{\min}$ and MMD .

Bibliography

- Barducci, A., and Pippi, I., 1996. Temperature and emissivity retrieval from remotely sensed images using the “grey body emissivity” method. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(3), 681–695.
- Becker, F., and Li, Z. L., 1990. Temperature-independent spectral indices in thermal infrared bands. *Remote Sensing of Environment*, **32**, 17–33.
- Clark, R. N., Swayze, G. A., Livo, K. E., Kokaly, R. F., Sutley, S. J., Dalton, J. B., McDougal, R. R., and Gent, C. A., 2003. Imaging spectroscopy: earth and planetary remote sensing with the USGS Tetracorder and expert systems. *Journal of Geophysical Research*, doi:10.1029/2002JE001847.
- Dash, P., 2005. *Land Surface Temperature and Emissivity Retrieval from satellite Measurements*. Dissertation, Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft, Wissenschaftliche Berichte, FZKA 7095, 99 pp. Available from <http://bibliothek.fzk.de/zb/berichte/FZKA7095.pdf>. Last Accessed July 7, 2013.
- Gillespie, A. R., 1985. Lithologic mapping of silicate rocks using TIMS. In *The TIMS Data Users' Workshop*, June 18–19, Jet Propulsion Laboratory Publication 86–38. Pasadena, CA: Jet Propulsion Lab, pp. 29–44.
- Gillespie, A. R., Kahle, A. B., and Walker, R. E., 1986. Color enhancement of highly correlated images I. Decorrelation and HSI contrast stretches. *Remote Sensing of Environment*, **20**, 209–235.
- Gillespie, A. R., Matsunaga, T., Rokugawa, S., and Hook, S. J., 1998. Temperature and emissivity separation from advanced spaceborne thermal emission and reflection radiometer (ASTER) images. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 1113–1126.
- Gustafson, W. T., Gillespie, A. R., and Yamada, G., 2006. Revisions to the ASTER temperature/emissivity separation algorithm. In Sobrino, J. A. (ed.), *Second Recent Advances in Quantitative Remote Sensing*. Spain: Publicacions de la Universitat de València, pp. 770–775, ISBN 84-370-6533-X; 978-84-370-6533-5.
- Hackwell, J. A., Warren, D. W., Bongiovanni, R. P., Hansel, S. J., Hayhurst, T. L., Mabry, D. J., Sivjee, M. G., and Skinner, J. W., 1996. LWIR/MWIR imaging hyperspectral sensor for airborne and ground-based remote sensing. *Proceedings-SPIE The International Society For Optical Engineering*, **2819**, 102–107.
- Hook, S. J., Gabell, A. R., Green, A. A., and Kealy, P. S., 1992. A comparison of techniques for extracting emissivity information from thermal infrared data for geologic studies. *Remote Sensing of Environment*, **42**, 123–135.
- Hunt, G., 1980. Electromagnetic radiation: the communication link in remote sensing. In Siegal, B. S., and Gillespie, A. R. (eds.), *Remote Sensing in Geology*. New York: Wiley, pp. 5–45.
- Jaggi, S., Quattrochi, D., and Baskin, R., 1992. An algorithm for the estimation of bounds on the emissivity and temperatures from thermal multispectral airborne remotely sensed data (Abstract). In Realmuto, V. J. (ed.), *Summary of the Third Annual JPL Airborne Geoscience Workshop*, June 1–5, Jet Propulsion Laboratory Publication 92–14. Pasadena, CA: Jet Propulsion Lab, pp. 22–24.
- Kahle, A. B., and Rowan, L. C., 1980. Evaluation of multispectral middle infrared aircraft images for lithological mapping in the east Tintic Mountains, Utah. *Geology*, **8**, 234–239.
- Kahle, A. B., Madura, D. P., and Soha, J. M., 1980. Middle infrared multispectral aircraft scanner data: analysis for geological applications. *Applied Optics*, **19**, 2279–2290.
- Kealy, P. S., and Gabell, A. R., 1990. Estimation of emissivity and temperature using alpha coefficients. In *Proceedings of the 2nd TIMS Workshop*, Jet Propulsion Laboratory Publication 90–55. Pasadena, CA: Jet Propulsion Lab, pp. 11–15.
- Kealy, P. S., and Hook, S. J., 1993. Separating temperature and emissivity in thermal infrared multispectral scanner data: implication for recovering land surface temperatures. *IEEE Transactions on Geoscience and Remote Sensing*, **31**(6), 1155–1164.
- Liang, S. L., 1997. Retrieval of land surface temperature and water vapor content from AVHRR thermal imagery using artificial neural network. *International Geoscience and Remote Sensing Symposium Proceedings*, **3**, 1959–1961.
- Lyon, R. J. P., 1965. Analysis of rocks by spectral infrared emission (8 to 25 microns). *Economic Geology*, **60**, 715–736.
- Mao, K., Shi, J., Tang, H., Li, Z.-L., Wang, X., and Chen, K.-S., 2008. A neural network technique for separating land surface emissivity and temperature from ASTER imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(1), 200–208.
- Matsunaga, T., 1994. A temperature-emissivity separation method using an empirical relationship between the mean, the maximum, and the minimum of the thermal infrared emissivity spectrum. *Journal of the Remote Sensing Society of Japan*, **14**(2), 230–241 (in Japanese with English abstract).
- Mushkin, A., Balick, L. K., and Gillespie, A. R., 2005. Extending surface temperature and emissivity retrieval to the mid-infrared ($3\text{--}5 \mu\text{m}$) using the Multispectral Thermal Imager (MTI). *Remote Sensing of Environment*, **98**, 141–151.

- Norman, J. M., and Becker, F., 1995. Terminology in thermal infrared remote sensing of natural surfaces. *Remote Sensing Reviews*, **12**, 159–173.
- Palluconi, F. D., and Meeks, G. R., 1985. *Thermal Infrared Multispectral Scanner (TIMS): An Investigator's Guide to TIMS Data*. Jet Propulsion Laboratory Publication 85–32. Pasadena, CA: Jet Propulsion Lab, 14 pp.
- Realmuti, V. J., 1990. Separating the effects of temperature and emissivity: emissivity spectrum normalization. In *Proceedings of the 2nd TIMS Workshop*, Jet Propulsion Laboratory Publication 90–55. Pasadena, CA: Jet Propulsion Lab, pp. 31–36.
- Salisbury, J. W., and D'Aria, D., 1992. Emissivity of terrestrial materials in the 8–14 μm atmospheric window. *Remote Sensing of Environment*, **42**, 83–106.
- Slater, P. N., 1980. *Remote Sensing, Optics and Optical Systems*. Reading, MA: Addison–Wesley, p. 575.
- Soha, J. M., and Schwartz, A. A., 1978. Multispectral histogram normalization contrast enhancement. In *Proceedings of the 5th Canadian Symposium on Remote Sensing*, Victoria, British Columbia, Canada, pp. 86–93.
- Wan, Z., 1999. MODIS land-surface temperature algorithm theoretical basis document (LST ATBD), Version 3.3. *NASA Contract NAS5-31370*, 37 pp.
- Watson, K., 1992a. Two-temperature method for measuring emissivity. *Remote Sensing of Environment*, **42**, 117–121.
- Watson, K., 1992b. Spectral ratio method for measuring emissivity. *Remote Sensing of Environment*, **42**, 113–116.

Cross-references

Crop Stress
Fields and Radiation
Land Surface Temperature
Optical/Infrared, Radiative Transfer
Volcanism

LAND SURFACE ROUGHNESS

Thomas Farr
Jet Propulsion Laboratory, California Institute of
Technology, Pasadena, CA, USA

Synonyms

Microrelief; Microtopography

Definition

Surface roughness is usually defined at the human scales of centimeter to a few meter; larger scales are usually considered as topography. Relief at these scales is familiar to field geologists working at the outcrop scale and those interested in interpretation of landforms and earth-surface processes that form and modify them.

Scientific usefulness

One important surficial geologic process is aeolian erosion, transport, and deposition of sediments. The shear stress wind produces at the earth's surface is strongly affected by the surface roughness. The aerodynamic roughness

parameter, z_0 , depends on the wind speed profile as a function of height about the ground (Greeley et al., 1997). This parameter is used by geologists interested in aeolian processes as well as climatologists seeking to quantify atmospheric coupling with the solid earth.

Windblown dust and sand can also modify surface roughness by mantling and attenuating surface roughness (Farr, 1992; Arvidson et al., 1993). This can lead to estimates of relative age for surfaces such as lava flows or alluvial fans exposed to the same rate of aeolian deposition (Farr, 1992; Farr and Chadwick, 1996).

Streambed and ocean-bottom roughness also affect the flow and transport capabilities of water in those environments (e.g., Butler et al., 2001).

Other geologic processes produce or modify surface roughness, in particular volcanic eruptions which may mantle surfaces with ash or produce new roughness elements through extrusion of lava flows which can be relatively smooth pahoehoe or extremely rough aa. Roughness of lava flows can provide information on their eruption characteristics, such as rate and temperature (e.g., Lescinsky et al., 2007).

Land surface roughness strongly affects many remote sensing techniques. Observations of reflected visible-near-infrared wavelengths are affected by sub-resolution self-shadowing of roughness elements. Thus, rougher surfaces are darker, and the shadows are illuminated by sky light or reflections from adjacent land, shifting the spectral signature of the surface (Adams and Gillespie, 2006). At thermal infrared and microwave wavelengths, which are dominated by emission from solar-heated surfaces, roughness as well as larger-scale topography affects the initial heating of the surface while roughness also affects the efficiency of emission (Ulaby et al., 1982). Active microwave (radar) systems image surfaces through scattering of a transmitted wave from the surface. Smooth surfaces at the scale of the wavelengths, which are typically centimeter-meter, reflect energy away from the receiving antenna and are imaged as dark surfaces, while rough surfaces scatter the incident energy in all directions and show up in bright tones on radar images (Henderson and Lewis, 1998).

Much work has gone into quantitative models which seek to remove the effects of roughness on sub-resolution shadowing and thermal heating and emission (Tsang et al., 2000; Adams and Gillespie, 2006). In the radar area, inversion models have been developed which estimate the surface roughness from radar observations at different angles, polarizations, and wavelengths (Ulaby et al., 1982; Van Zyl et al., 1991; Evans et al., 1992; Dubois et al., 1995; Tsang et al., 2000).

Quantifying surface roughness

Good reviews of techniques for describing quantitatively surface roughness can be found in Dierking (1999), Thomas (1999), Shepard et al. (2001), and Campbell (2002), Chap. 3. The simplest description of surface

roughness is an estimate of the standard deviation (or root-mean square: RMS) of the surface heights (Table 1).

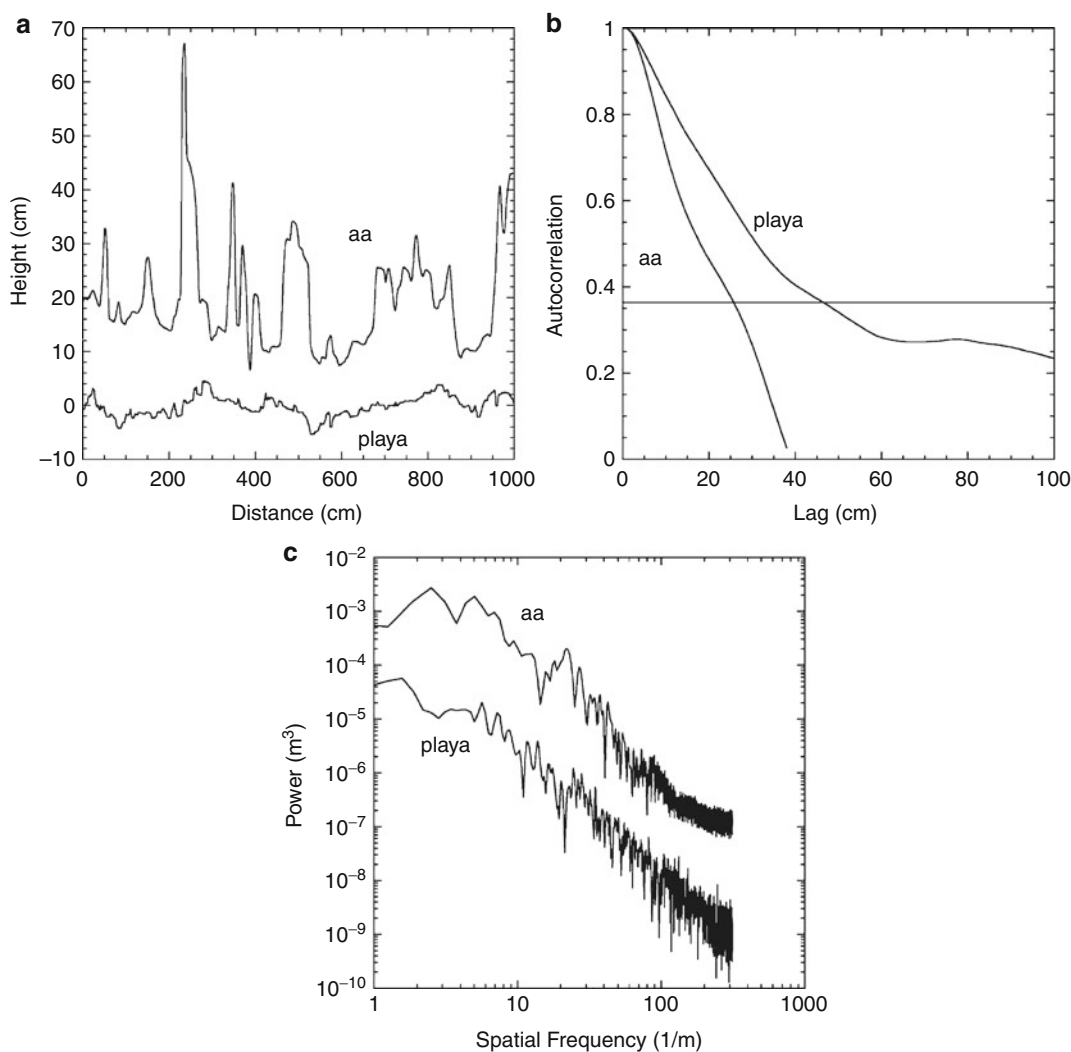
Describing the roughness of a surface by its RMS height leaves out any description of the scales of the

Land Surface Roughness, Table 1 Measures of surface roughness for two natural surfaces. aa is a rough lava flow surface at Pisgah lava field in the Mojave Desert. Playa is a smooth dry lake surface at Lunar Crater volcanic field in central Nevada. Profiles were measured at 1 cm spacing

		aa	Playa
RMS height (cm)		9.8	1.9
Correlation length (cm)		26	46
Power spectrum	Slope	-2.19	-2.24
	Offset	-1.69	-3.45

roughness. One way to describe the scale of the roughness is to calculate the correlation length of profiles. Correlation length is a measure of how quickly heights change when moving along a profile. The autocorrelation function for a surface profile is calculated by sequentially stepping the profile across a stationary copy, multiplying, and normalizing. The autocorrelation is unity for 0 steps, or lags, and then drops as the number of lags increases (Figure 1). The rate of the drop-off, measured by the lag at which the autocorrelation value drops to $1/e$, is called the correlation length, l . Smoother surfaces tend to have larger correlation lengths (Ulaby et al., 1982; p. 822).

Another way to describe quantitatively both the amplitude and scale of surface roughness is through the power spectrum, or power spectral density, usually of profiles. The power spectrum is basically the Fourier



Land Surface Roughness, Figure 1 Profiles and surface roughness measures of two natural surfaces: an aa lava surface (*rough*) and a playa surface (*smooth*). (a) Profiles: aa profile has been offset 20 cm for clarity. (b) Autocorrelation functions for the two profiles. Note rapid drop-off of aa autocorrelation. Horizontal line is at $1/e$. (c) Power spectra of the two profiles. Note they are nearly linear (in log-log plot) and parallel. Playa has much less power at all spatial frequencies.

transform of the profile (or two-dimensional microtopography) (e.g., Bendat and Piersol, 1986; Brown and Scholz, 1985; Austin et al., 1994). This produces a plot showing power, or variance, as a function of spatial frequency or scale (Figure 1). When plotted in log-log coordinates, the functions are found to be approximately linear (e.g., Berry and Hannay, 1978; Farr, 1992; Shepard et al., 2001), indicating a power-law relationship between roughness and scale. This relationship simplifies the quantitative description of the power spectrum of a profile to two numbers: the slope and offset (Table 1). Power spectrum slope is a measure of self-similarity and is related to fractal dimension of a profile, D , by (Brown, 1985):

$$D = \frac{5 + \text{slope}}{2}$$

Power spectrum offset is a measure of overall roughness, sometimes called “roughness amplitude” (Huang and Turcotte, 1989, 1990; Goff, 1990).

Measurement of surface roughness

Measurement of surface roughness at scales of centimeter to several meters is difficult, especially as the area covered must be large enough to make statistically significant calculations of quantities described above for natural surfaces. This usually means measuring an area or profiles 10–20 m or more in size. Techniques used to date include templates (Shepard et al., 2001); stereophotogrammetry from handheld and balloon-borne cameras and from a helicopter (Wall et al., 1991; Farr, 1992); and more recently ground-based lidar systems (Morris et al., 2008). All of these techniques have provided cm or better resolution, but helicopter stereophotogrammetry and ground-based lidar have provided the best coverage.

Summary

Land surface roughness at scales of centimeter to several meters is important in several areas of earth science as well as in the interpretation of remote sensing data. Roughness can be quantified in a variety of ways, but power spectral analysis is best at describing roughness and its scaling properties.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Adams, J. B., and Gillespie, A. R., 2006. *Remote Sensing of Landscapes with Spectral Images: a Physical Modeling Approach*. Cambridge, UK: Cambridge University Press, p. 362.
- Arvidson, R. E., Shepard, M. K., Guinness, E. A., Petroy, S. B., Plaut, J. J., Evans, D. L., Farr, T. G., Greeley, R., Lancaster, N., and Gaddis, L. R., 1993. Characterization of lava-flow degradation in the Pissgah and Cima volcanic fields, California, using Landsat Thematic Mapper and AIRSAR data. *Geological Society of America Bulletin*, **105**, 175–188.
- Austin, R. T., England, A. W., and Wakefield, G. H., 1994. Special problems in the estimation of power-law spectra as applied to topographical modeling. *IEEE Transactions on Geoscience and Remote Sensing*, **32**, 928–939.
- Bendat, J. S., and Piersol, A. G., 1986. *Random Data, Analysis and Measurement Procedures*, 2nd edn. New York: Wiley, p. 566.
- Berry, M. V., and Hannay, J. H., 1978. Topography of random surfaces, comment and reply. *Nature*, **273**, 573.
- Brown, S. R., 1985. A note on the description of surface roughness using fractal dimension. *Geophysical Research Letters*, **14**, 1095–1098.
- Brown, S. R., and Scholz, C. H., 1985. Broad bandwidth study of the topography of natural rock surfaces. *Journal of Geophysical Research*, **90**, 12575–12582.
- Butler, J. B., Lane, S. N., and Chandler, J. H., 2001. Characterization of the structure of river-bed gravels using two-dimensional fractal analysis. *Mathematical Geology*, **33**, 301–330.
- Campbell, B. A., 2002. *Radar Remote Sensing of Planetary Surfaces*. Cambridge, UK: Cambridge University Press, p. 331.
- Dierking, W., 1999. Quantitative roughness characterization of geological surfaces and implications for radar signature analysis. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 2397–2412.
- Dubois, P. C., van Zyl, J., and Engman, T., 1995. Measuring soil moisture with imaging radar. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 915–926.
- Evans, D. L., Farr, T. G., and van Zyl, J. J., 1992. Estimates of surface roughness derived from synthetic aperture radar (SAR) data. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, 382–389.
- Farr, T. G., 1992. Microtopographic evolution of lava flows at Cima volcanic field, Mojave Desert, California. *Journal of Geophysical Research*, **97**, 15171–15179.
- Farr, T. G., and Chadwick, O. A., 1996. Geomorphic processes and remote sensing signatures of alluvial fans in the Kun Lun Mountains, China. *Journal of Geophysical Research*, **101**, 23091–23100.
- Goff, J. A., 1990. Comment on “Fractal mapping of digitized images: application to the topography of Arizona and comparisons with synthetic images”. *Journal of Geophysical Research*, **95**, 5159–5161.
- Greeley, R., Blumberg, D., McHone, J. F., Dobrovolskis, A., Iversen, J. D., Lancaster, N., Rasmussen, K. R., Wall, S. D., and White, B. R., 1997. Application of spaceborne radar laboratory data to the study of aeolian processes. *Journal of Geophysical Research*, **102**, 10,971–10,983.
- Henderson, F. M., and Lewis, A. J. (eds.), 1998. *Principles and Applications of Imaging Radar, Manual of Remote Sensing*. New York: Wiley, Vol. 2, p. 866.
- Huang, J., and Turcotte, D. L., 1989. Fractal mapping of digitized images: application to the topography of Arizona and comparisons with synthetic images. *Journal of Geophysical Research*, **94**, 7491–7495.
- Huang, J., and Turcotte, D. L., 1990. Fractal image analysis: application to the topography of Oregon and synthetic images. *Journal of the Optical Society of America*, **7**, 1124–1130.
- Lescinsky, D. T., Skoblenick, S. V., and Mansinha, L., 2007. Automated identification of lava flow structures using local Fourier spectrum of digital elevation data. *Journal of Geophysical Research*, **112**, doi:10.1029/2006JB004263.
- Morris, A. R., Anderson, F. S., Mougini-Mark, P. J., Haldemann, A. F. C., Brooks, B. A., and Foster, J., 2008. The roughness of Hawaiian volcanic terrains. *Journal of Geophysical Research*, **113**, E12007.

- Shepard, M. K., Campbell, B. A., Bulmer, M. H., Farr, T. G., Gaddis, L. R., and Plaut, J. J., 2001. The roughness of natural terrain: a planetary and remote sensing perspective. *Journal of Geophysical Research*, **106**, 32,777–32,795.
- Thomas, T. R., 1999. *Rough Surfaces*. London: Imperial College London, p. 278.
- Tsang, L., Kong, J. A., and Ding, K. H., 2000. *Scattering of Electromagnetic Waves: Theories and Applications*. New York: Wiley, p. 426.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1982. *Microwave Remote Sensing*. Reading: Addison-Wesley, Vol. 1, 2, 3, p. 2162.
- van Zyl, J. J., Burnette, C. F., and Farr, T. G., 1991. Inference of surface power spectra from inversion of multifrequency polarimetric radar data. *Geophysical Research Letters*, **18**, 1787–1790.
- Wall, S. D., Farr, T. G., Muller, J.-P., Lewis, P., and Leberl, F. W., 1991. Measurement of surface microtopography. *Photogrammetric Engineering and Remote Sensing*, **57**, 1075–1078.

Cross-references

[Geomorphology](#)
[Geophysical Retrieval, Inverse Problems in Remote Sensing](#)
[Land Surface Emissivity](#)
[Lidar Systems](#)
[Microwave Radiometers](#)
[Microwave Surface Scattering and Emission](#)
[Radars](#)
[Radar, Scatterometers](#)
[Radar, Synthetic Aperture](#)
[Surface Truth](#)
[Trafficability of Desert Terrains](#)

LAND SURFACE TEMPERATURE

Alan Gillespie
 Department of Earth and Space Sciences, University of
 Washington, Seattle, WA, USA

Definition

Land surface temperature (LST). Average temperature of an element of the exact surface of the Earth calculated from measured radiance (for a complete definition, see Norman and Becker, 1995).

Blackbody. An ideal material absorbing all incident energy or emitting all thermal energy possible. A cavity with a pinhole aperture approximates a blackbody.

Color temperature. Temperature satisfying Planck's law for spectral radiances measured at two different wavelengths: for a gray body (this entry), for any emitter, or the blackbody temperature for which visual color is the same as some other source (e.g., in photography).

Emissivity ϵ . The efficiency with which a surface radiates its thermal energy.

Irradiance. The power incident on a unit area, integrated over all directions (W m^{-2}).

Longwave infrared (LWIR). For most terrestrial surfaces (340–240 K) peak thermal emittance occurs at LWIR (8–14 μm).

Mid-infrared (MIR). Forest fires (1,000–600 K) have peak thermal emittances in the MIR (3–5 μm).

MODTRAN. A computer code package that describes the generation and transmission of energy in the Earth's atmosphere. MODTRAN can be used to predict τ , $S_{\uparrow}$, and $I_{\downarrow}$ for terrestrial remote sensing (Berk et al., 2005).

Path radiance $S_{\uparrow}$. The power per unit area incident on a detector and emitted upward from within the atmosphere ($\text{W m}^{-2} \text{ sr}^{-1}$).

Radiance. The power per unit area from a surface directed toward a sensor, in units of $\text{W m}^{-2} \text{ sr}^{-1}$.

Reflectivity ρ . The efficiency with which a surface reflects energy incident on it.

Shortwave infrared (SWIR). Erupting basaltic lavas (1,400 K) have their maximum thermal emittance at 2.1 μm in a third atmospheric window (0.4–2.5 μm). Part of this spectral region (1.4–2.5 μm) is called the SWIR.

Sky irradiance $I_{\downarrow}$. The irradiance on the Earth's surface originating as thermal energy radiated downward by the atmosphere (W m^{-2}).

Spectral radiance L . Radiance per wavelength, in units of $\text{W m}^{-2} \mu\text{m}^{-1} \text{ sr}^{-1}$.

Thermal infrared (TIR). Thermal energy is radiated from a body at frequencies or wavelengths in proportion to its temperature. The wavelengths for which this radiant energy is significant for most terrestrial surfaces (1.4–14 μm) are longer than the wavelength of visible red light and hence are known as thermal infrared. The TIR is subdivided into three ranges (LWIR, MIR, and SWIR) for which the atmosphere is transparent (atmospheric “windows”) so that the energy can be measured from space.

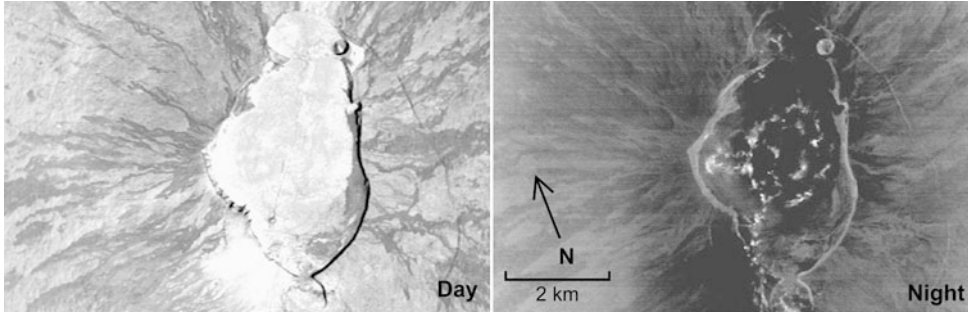
Transmissivity τ . The efficiency with which a material transmits energy incident on it.

Introduction

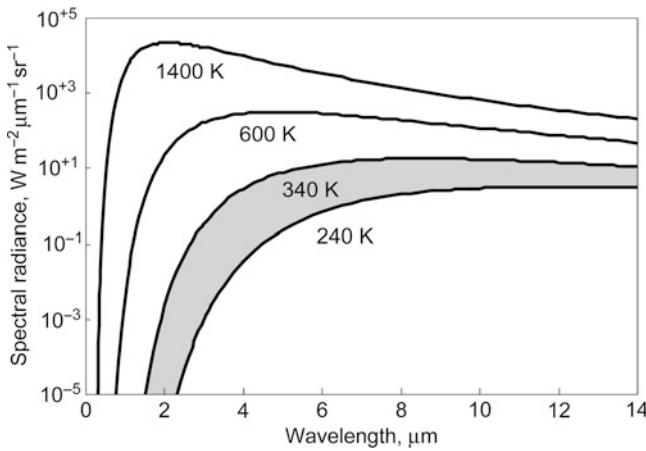
Temperature is a fundamental property of the Earth's surface that can be determined remotely. At all spatial scales, temperature is used in energy-balance and ecological studies, and it is important for geothermal and volcanic monitoring. At fine scales, temperature images are used to identify “hot spots” in fighting fires. Recently, they have been important for monitoring the seasonal onset of melting and freezing conditions in the Arctic, where melting is anticipated to occur earlier, and freezing later, as “global warming” worsens.

The land surface temperature (LST) itself is not measured directly from airborne or spaceborne sensors. Instead, the surface radiates thermal energy in the thermal infrared (TIR) part of the spectrum. The energy is radiated in proportion to temperature; therefore, the temperature can be calculated from it.

Temperature is variable at all temporal scales, but diurnal and seasonal fluctuations are perhaps the most noticeable. Figure 1 compares day-and nighttime TIR images of the same scene, the summit of Mauna Loa, Hawaii. The daytime image is dominated by cooling due



Land Surface Temperature, Figure 1 TIR radiance images of the caldera of Mauna Loa volcano, Hawaii. Satellite images from the multispectral thermal imager (MTI), acquired at $8.65 \mu\text{m}$ 10/09/00 (day) and 10/12/00 (night).



Land Surface Temperature, Figure 2 The spectrum of emitted spectral radiance for surfaces at different temperatures (note logarithmic scale on y-axis). Most terrestrial remote sensing is for surfaces between 340 and 240 K (shaded area). The peak spectral radiance occurs at longer wavelengths for cooler surfaces, but at any given wavelength it rises exponentially with temperature.

to topographic shading and thermal inertia of the lavas. Darker surfaces absorb more sunlight and so will be warmer. At night, the topographic effects are minimal, and white fractures in the lava (70°C) delineate zones of geothermal heat rising within the caldera.

The basic equation governing relationship of surface or “skin” temperature and radiant energy is Planck’s law (Figure 2):

$$B(\lambda, T) = \frac{c_1}{\pi \lambda^5} \left(\exp\left(\frac{c_2}{\lambda T}\right) - 1 \right)^{-1} \quad (1)$$

where B is the spectral radiance ($\text{Wm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) at wavelength λ (μm) and temperature T (K) and $c_1 = 3.7418 \cdot 10^{-16} \text{ Wm}^2$ and $c_2 = 14,388 \mu\text{m K}$. It is assumed that the surface emits equally in all directions.

Integrating Equation 1 across all wavelengths gives the Stefan-Boltzmann law

$$j = \frac{\sigma}{\pi} T^4 \quad (2)$$

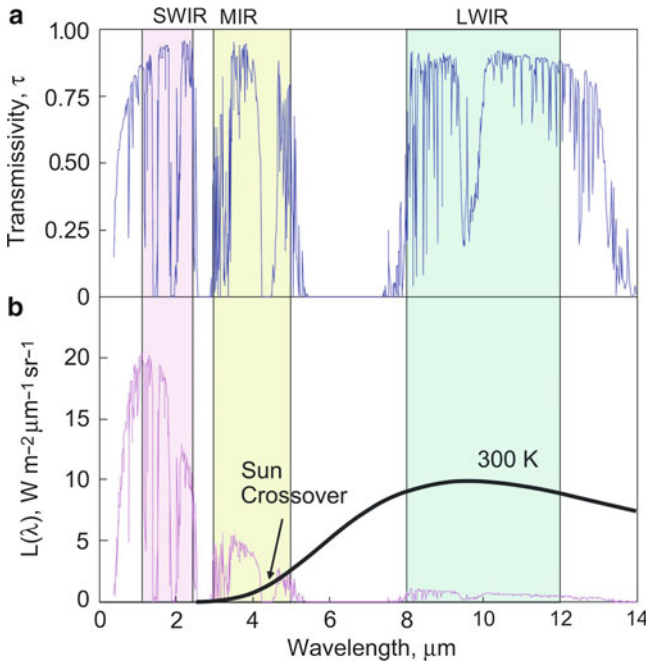
where j is the radiance ($\text{Wm}^{-2} \text{sr}^{-1}$) and $\sigma = 5.669 \times 10^8 \text{ Wm}^{-2} \text{K}^{-4}$. For broadband measurements of j , inversion of Equation 2 could yield the LST.

Most surfaces are not perfect emitters, however, and a parameter called the emissivity (ε) is used to describe the efficiency with which they radiated heat. For a perfect emitter, $\varepsilon = 1$, but commonly $0.8 < \varepsilon < 1.0$. Therefore, for most surfaces $j = \varepsilon \sigma \pi^{-1} T^4$. Furthermore, ε varies with wavelength for many surfaces, including most geological surfaces of rock and soil (e.g., Lyon, 1965), so that Equation 1 also must be modified by including the multiplicative factor, $\varepsilon(\lambda)$. Inversion of this modified Planck function is generally used to calculate LST. Because it is underdetermined, however, additional information is required for solution. Providing these constraints is not straightforward, which is why there are numerous approaches. Figure 3 of “Land Surface Emissivity,” compares radiance, temperature, and emissivity data for a TIR image.

The discussion that follows addresses several factors that complicate the calculation of LST: effects of atmospheric absorption, emission, and scattering on the measured spectral radiance; effects of land surface heterogeneity, including skin effects, and spatial integration during measurement of spectral radiance; short-term temporal fluctuations in the temperature field; and simplifying assumptions and algorithms used to recover temperatures.

Atmospheric effects

The Earth’s atmosphere interacts with surface-emitted radiation as a function of surface elevation and atmospheric conditions, especially temperature and humidity profiles. Constituent gases such as ozone are also important. Atmospheric gases absorb differently across the spectrum, and at some wavelengths it is almost opaque. Remote sensing of the surface can only occur in atmospheric windows in spectral regions between these opaque bands. The three main windows for TIR remote



Land Surface Temperature, Figure 3 Radiance and transmissivity spectra. (a) Atmospheric transmissivity at sea level, zenith-nadir path (MODTRAN, mid-latitude summer atmosphere: Berk et al., 2005). (b) 300 K blackbody and sunlight reflected from a sea-level horizontal surface with 10 % albedo. Most remote-sensing detectors operate only within transparent atmospheric “windows” shown as labeled and shaded bars.

sensing are the longwave, mid-, and shortwave infrared (Figure 3). The maximum transmissivity $\tau(\lambda)$ of the atmosphere in the TIR is about 90 % (sea level). Therefore, it is necessary to account for atmospheric absorption and other effects quantitatively if accurate LSTs are to be recovered.

In addition to absorbing surface-emitted radiance, the atmosphere emits its own thermal energy in all directions. Two summary terms are important in the radiative transfer of energy from Earth to satellite: the upwelling “path radiance” ($L_{\uparrow}(\lambda)$) that never interacts with the land surface and the down-welling “sky irradiance” ($I_{\downarrow}(\lambda) = \pi L_{\downarrow}(\lambda)$), where $L_{\downarrow}$ is the equivalent average directional radiance (per steradian) term. Using these factors, the modified Planck function can be rewritten as

$$L(\lambda, T) = \tau(\lambda)\varepsilon(\lambda)B(\lambda, T) + L_{\uparrow}(\lambda) + \tau(1 - \varepsilon(\lambda))L_{\downarrow}(\lambda) \quad (3)$$

where L is the spectral radiance measured at the satellite and where the reflectivity and emissivity of the land surface are related by Kirchhoff’s law:

$$\rho(\lambda) = (1 - \varepsilon(\lambda)). \quad (4)$$

The atmospheric terms and τ , $L_{\uparrow}$, and $L_{\downarrow}$ are themselves functions of the angle of the exitant ray relative to local

zenith, as well as of the factors mentioned above. In evaluating Equation 3, it is important to note that there are five unknowns for a single spectral radiance measured from a homogeneous, isothermal surface. For multispectral data with n image channels at different wavelengths, there are $(1 + 4n)$ unknowns per measurement. This indeterminacy must be eliminated if LST is to be recovered.

Compensation for atmospheric effects

Before compensation can be attempted, the atmosphere must be characterized and the atmospheric parameters τ , $L_{\uparrow}$, and $L_{\downarrow}$ must be estimated. Probably the most common approach has been to use generalized atmospheric profiles of pressure, temperature, and humidity, from which the three parameters are estimated via an atmospheric radiative transfer model such as MODTRAN (e.g., Berk et al., 2005). Such an approach has been used in the retrieval of LST from Landsat TM5, for example, (Sobrino et al., 2004). This approach requires accounting for ground elevations using a digital elevation model (DEM).

It is potentially more accurate to estimate τ , $L_{\uparrow}$, and $L_{\downarrow}$ pixel by pixel, from the remote-sensing data themselves, as is done by MODIS to recover atmospheric profiles of temperature and humidity (Seaman et al., 2006). These profiles are then used in atmospheric radiative transfer models to calculate τ , $L_{\uparrow}$ and $L_{\downarrow}$. But profile recovery requires a complex sensor with many bands in and out of the LWIR window, and most remote-sensing platforms do not support this level of TIR measurement. Some airborne scanners such as NASA’s MODIS Airborne Simulator (MAS) do allow measurements across an H_2O absorption band at $0.93 \mu m$, from which total column water (TCW) can be estimated (Conel et al., 1988; Gao and Goetz, 1990). TCW can be used to constrain MODTRAN by scaling the default or assumed humidity profiles such that the measured integral is the TCW (Tonooka et al., 2005).

ASTER does not use internal image data to estimate atmospheric characteristics, but relies on reanalysis data from radiosonde balloons, launched routinely from weather stations and airports, interpolated to a 1° latitude/longitude grid and sampled every 6 h (e.g., Kalnay et al., 1996; Tonooka and Palluconi, 2005). They and a DEM are input to MODTRAN. The atmospheric data are available over the Internet (<http://www.cpc.ncep.noaa.gov/products/wesley/reanalysis.html>, last accessed 15 August 2008).

Because τ , $L_{\uparrow}$, and $L_{\downarrow}$ are partly correlated, it has proven possible to use just two LWIR channels in order to estimate them, provided the surface emissivities are known (Wan and Dozier, 1996). This is the case for the oceans, and such a two-channel or “split-window” algorithm for calculating sea surface temperature (SST) was developed and tested successfully early in the history of terrestrial remote sensing (Anding and Kauth, 1970). Details of this approach for SSTs are available in “Ocean, Measurements and Applications” and “Sea Surface Temperature.”

For images of constant or known high emissivity (~ 1.0), especially of grasslands or forests, it is possible to normalize the atmospheric τ and $L_{\uparrow}$ from channel to channel of a multispectral TIR image (Young et al., 2002), eliminating the need to measure it independently. In this approach, $L_{\downarrow}$ is overlooked because ρ is near zero. A reference channel ref is chosen, and variation diagrams are constructed in which $L(\lambda_i)$ is plotted against $L(\lambda_{ref})$ for enough pixels that a range of LSTs are represented. A line fit to the high radiance edge of the data cluster will have a slope of τ and an intercept of $L_{\downarrow}$, where these parameters are relative to those for the reference channel. This approach is successful in removing the appearance of atmospheric effects from multispectral images, but a whole-spectrum scaling still needs to be applied in order to accurately portray land-leaving spectral radiance.

Land surface heterogeneity

The temperature of the Earth's surface is variable in space and time at a wide range of scales. Both types of variability must be considered in TIR remote sensing.

Energy absorbed or radiated from the surface changes the "skin temperature" of the immediate surface. This establishes a temperature gradient into the surface, and the skin temperature is not a completely accurate representation of the temperature of even the upper few centimeters of the Earth. Sensible heat transport, for example, cooling by wind, also creates a temperature gradient. Because wind is rarely constant, the LST fluctuates on a scale of seconds or minutes, with an amplitude of a few Kelvin, especially over vegetation. The ephemeral variability of LST means that measurements by remote scanners, with a dwell time of milliseconds at each pixel, are hard to compare to other remote or in situ measurements, since the exact time of measurements are unlikely to be the same. This is sometimes called the "snapshot" problem.

These considerations are especially important for energy-balance studies and validation exercises for LST products. However, rapid temporal fluctuations are most severe at fine scales (e.g., 10^0 – 10^1 m) and are reduced at the moderate scales (e.g., 10^1 – 10^3 m) of many civilian TIR scanners such as MODIS or GOES.

LST variability from point to point also poses a problem, because the measured radiance is integrated from surface elements of different temperatures:

$$L(\lambda, T_e) = \int B(\lambda, T_{x,y}) dx dy. \quad (5)$$

where (x,y) are geographic coordinates. The temperature recovered by inversion of Equation 1 for the integrated spectral radiance yields an effective temperature, T_e . Because radiance is an exponential function of temperature, the exact relationship of T_e and the actual distribution of $T_{x,y}$ are hard to predict exactly and change over the day. For example, in a scene half covered by trees,

T_e in the morning will be weighted toward the temperature of the trees, which warm faster than rock, but in the evening T_e will be weighted toward the temperature of the rock surface, which cools more slowly. This may not be a serious concern to all analysts. As an example, for an afternoon scene with 50 % trees at 280 K and 50 % rocks at 310 K, the effective temperature will be 296.3 K, only 1.3 K greater than the average skin temperature. However, this error is noticeable in, for example, an ASTER image.

Directional effects

Many remotely sensed images are acquired looking straight down through the atmosphere, but wide-angle images from airborne scanners and from AVHRR and MODIS have off-nadir look angles as large as $\sim 55^\circ$. In addition to the dependency of atmospheric effects on the slant range, there are three effects that must be considered in calculating LST. These can lead to gradients in images of LST, such that the temperatures on the edges of the image may be hard to relate to those for equivalent surfaces viewed at nadir. First, in viewing rough surfaces at different angles, some views will include more shaded (typically cooler) surfaces, and others will include more sunlit surfaces (typically warmer). The magnitude of this effect depends on the time of day and the position of the sun, and the effect is minimized at night. However, during the day, the temperature difference between sunlit and shadowed surfaces may be 10 K or more, and the effect may be a few Kelvin across an LST image. This effect is not artifactual in that it is a faithful representation of the T_e field, and to remove it requires careful modeling and measurement of roughness. For daytime data, the ASTER stereo capability permits optical assessments of roughness (Mushkin and Gillespie, 2005) that could serve as input for such models. MODIS LST standard products have had to be made with the effects of non-Lambertian surfaces viewed at a range of angles from nadir in mind (Wan, 1999).

Second, specular reflection (Snyder and Wan, 1998) is also a factor at oblique view angles. This reflection is most obvious for smooth, mirrorlike surfaces such as lakes and ponds. The reflected emittance may be from the sky ($L_{\downarrow}$) or from neighboring landscape elements. For images acquired within $\sim 42^\circ$ of nadir, reflected radiance is generally minimal.

Third, the bidirectional reflectance distribution (BRDF) of the surface may not be isotropic (e.g., Wan and Dozier, 1992). According to Kirchhoff's law, this means that a surface will look hotter from some angles than others. For rough surfaces, temperature differences due to shadowing (see above) can masquerade as an anisotropic BRDF; other surfaces may inherently be anisotropic. Because it is necessary to characterize the surface before the BRDF can be compensated for, it has been necessary to use generalized approaches (e.g., Wan, 1999).

Algorithms for calculating LST

Consideration of all the fluxes and factors contributing to the spatially integrated measurement of spectral radiance shows that accurate LST estimation is only possible under restricted circumstances or with independent constraints and simplifying assumptions. For example, it may be necessary to regard the landscape pixel as having a known, single emissivity ε and an effective temperature T_e which may, in fact, exist nowhere within the pixel. However, over decades of study, the assumptions have proven reasonable, and T_e has proven to be a useful measure.

Below, several common approaches for recovery of T_e are discussed. They are based either on the amplitude of the spectral radiance or on the shape of the spectrum if two or more image channels are measured. All approaches here are described assuming that the spectral range of measurement, $\Delta\lambda = f_2 - f_1$, is very narrow. If this is not the case, a calibration must be done to estimate the spectral radiance at the central wavelength of the sensor response $f(\lambda)$:

$$L(\lambda, T) \approx \frac{\int_{\lambda=f_1}^{f_2} f(\lambda)\varepsilon(\lambda)B(\lambda, T)d\lambda}{\int_{\lambda=f_1}^{f_2} f(\lambda)d\lambda} \quad (6)$$

Brightness temperature, T_b

The simplest estimate of LST is T_b , the temperature at which a blackbody ($\varepsilon \equiv 1.0$) would emit the remotely measured radiance. This approach is used when there is a single image channel. Finding T_b requires inverting the Planck function:

$$T_b = \frac{c_2}{\lambda \ln\left(\frac{c_1}{B(\lambda, T)\pi\lambda^5} + 1\right)} \quad (7)$$

where it is understood that λ refers to the central wavelength of the image channel.

Model temperature, T_m

In a slight variation, the surface emissivity may be estimated a priori to provide a refined estimate of LST (T_m). If the surface composition is known, the emissivity may be found from one of several spectral libraries available online (e.g., <http://speclib.jpl.nasa.gov>, last accessed 15 August 2008). For water and closed-canopy vegetation, the LWIR emissivity is >0.98 . Many soils also have a high emissivity, >0.95 . Lyon (1965) pointed out that, for most rocks, the emissivity at $\sim 10 \mu\text{m}$ is ~ 0.95 .

Once ε has been estimated, LST is found by

$$T_m = \frac{c_2}{\lambda \ln\left(\frac{c_1}{\varepsilon(\lambda)B(\lambda, T)\pi\lambda^5} + 1\right)} \quad (8)$$

Image classification has been used to determine surface composition, from which ε can be inferred. For example, van de Griend and Owe, (1993) showed that in the 8 – 14 μm spectral range, ε is highly correlated with NDVI, the normalized difference vegetation index, for different vegetation types:

$$\varepsilon \approx a + b \ln NDVI; NDVI = \frac{\rho_{0.85} - \rho_{0.65}}{\rho_{0.85} + \rho_{0.65}} \quad (9)$$

where ρ is evaluated at near-infrared ($\lambda = 0.85 \mu\text{m}$) and visible red ($\lambda = 0.65 \mu\text{m}$) wavelengths. Equation 9 allows ε to be estimated pixel by pixel, improving LST estimates especially for agricultural areas with simple geological substrates (homogeneous soils). Empirical coefficients a and b in Equation 9 are dependent on substrate emissivities, but if these can be estimated it is possible to extend greatly the range of surfaces for which accurate model LSTs can be calculated.

Model temperatures have been calculated for lava flowing from volcanoes, using Landsat Thematic Mapper (TM) Band 7 (2.25 μm) images and emissivities appropriate for lava (Pieri et al., 1990). The SWIR bands are effective because the peak thermal radiance is near 2 μm , whereas the emitted TIR radiance is too high and exceeds the dynamic range of the TIR band. Care must be taken because the emissivity of the hot lava appears to vary with temperature (Abtahi et al., 2001).

Color temperature, T_c

The simplest approach that utilizes the changing shape of the Planck function to estimate LST is a ratio of two spectral bands centered at λ_a and λ_b , and for which emissivities are known. The ratio is a monotonic function of temperature:

$$\frac{L(\lambda_a, T)}{L(\lambda_b, T)} = \frac{\varepsilon(\lambda_a)\lambda_b^5}{\varepsilon(\lambda_b)\lambda_a^5} \left(\frac{\exp\left(c_2(\lambda_b T)^{-1}\right) - 1}{\exp\left(c_2(\lambda_a T)^{-1}\right) - 1} \right). \quad (10)$$

Generally, it is assumed that $\varepsilon(\lambda_a) = \varepsilon(\lambda_b)$. Equation 10 is sometimes simplified further (Wien's approximation) by ignoring the (-1) term, such that

$$T_c = \left(\frac{c_2}{\lambda_b} - \frac{c_2}{\lambda_a} \right) \left(\ln \left(\frac{L(\lambda_a, T)\lambda_a^5}{L(\lambda_b, T)\lambda_b^5} \right) \right)^{-1} \quad (11)$$

For typical LSTs (e.g., 300 K), T_c is best calculated for widely separated central wavelengths, for example, 3 and 10 μm , because the ratio is more sensitive to LST, and therefore measurement precision has less of an effect on the recovered T_c . However, the use of MIR channels is generally limited to night time data because of the need to make large corrections for reflected sunlight.

Generalized split-window algorithm

Wan and Dozier, (1996) showed that provided the emissivity was known, the SST algorithm could be generalized for use over land. This algorithm is based on spectral radiance differences rather than ratios and has empirical coefficients a and b that permit the elimination of atmospheric effects. The form is

$$T_s = T_4 + \frac{1}{a-1}(T_4 - T_5) - \frac{b}{a-1} \quad (12)$$

T_4 and T_5 are the brightness temperatures for AVHRR (Advanced Very High Resolution Radiometer) bands 4 and 5 (10.8 and 11.9 μm) or MODIS bands 31 and 32. Equation 12 can be generalized further to account for directional effects, but is still of the form of a difference equation.

Summary and conclusions

Temperature and emissivity are generally both unknown for “geological” surfaces of rock and soil and must be solved for simultaneously using inversion of Planck’s equation. These algorithms thus go beyond just finding LST and are discussed in the entry “Land Surface Emissivity.”

Because of the underdetermined nature of the modified Planck equation (Equation 2), no single solution for LST has been found that satisfies all analysts. Therefore, literally dozens of algorithms have been proposed, tested, and applied. This entry summarized only a few, representative of the basic approaches.

Bibliography

- Abtahi, A. A., Kahle, A. B., Abbott, E. A., Gillespie, A. R., Sabol, D., Yamada, G., and Pieri, D., 2002. Emissivity changes in basalt cooling after eruption from Pu’u O’o, Kilauea, Hawaii. *Eos Transactions of the American Geophysical Union*, **83**(47), Fall Meeting Supplement, Abstract V71A–1263.
- Anding, D., and Kauth, R., 1970. Estimation of sea surface temperature from space. *Remote Sensing of Environment*, **1**, 217–220.
- Berk, A., Anderson, G. P., Acharya, P. K., Bernstein, L. S., Muratov, L., Lee, J., Fox, M., Adler-Golden, S. M., Chetwynd, J. H., Hoke, M. L., Lockwood, R. B., Gardner, J. A., Cooley, T. W., and Lewis, P. E., 2005. MODTRAN5: a reformulated atmospheric band model with auxiliary species and practical multiple scattering options. In *Proceedings of Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery*, SPIE, 5806: 662; DOI:10.1117/12.606026.
- Conel, J. E., Green, R. O., Carrere, V., Margolis, J. S., Alley, R. E., Vane, G., Bruegge, C. J., and Gary, B. L., 1988. Atmospheric water mapping with the airborne visible/infrared imaging spectrometer (AVIRIS), Mountain Pass, CA. In Vane, G. (ed.), *Proceedings of the AVIRIS Performance Evaluation Workshop*. JPL Publication 88–38, Pasadena: Jet Propulsion Lab., pp. 21–26.
- Gao, B.-C., and Goetz, A. F. H., 1990. Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. *Journal of Geophysical Research-Atmospheres*, **95**, 3549–3564.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebsuzaki, W., Higgins, W., Janowiak, J., Mo,

- K. C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R., and Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–470.
- Lyon, R. J. P., 1965. Analysis of rocks by spectral infrared emission (8 to 25 microns). *Economic Geology*, **60**(4), 715–736.
- Mushkin, A., and Gillespie, A. R., 2005. Estimating sub-pixel surface roughness using remotely sensed stereoscopic data. *Remote Sensing of Environment*, **99**, 75–83.
- Norman, J. M., and Becker, F., 1995. Terminology in thermal infrared remote sensing of natural surfaces. *Remote Sensing Reviews*, **12**, 159–173.
- Pieri, D. C., Glaze, L. S., and Abrams, M. J., 1990. Thermal radiance observations of an active lava flow during the June 1984 eruption of Mt. Etna. *Geology*, **18**, 1018–1022.
- Seaman, S. W., Borbas, E. E., Li, J., Menzel, W. P., and Gumley, L. E., 2006. MODIS atmospheric profile retrieval algorithm theoretical basis document, version 6. Available from http://modis.gsfc.nasa.gov/data/atbd/atbd_mod07.pdf. Accessed 7 June, 2013.
- Snyder, W. C., and Wan, Z.-M., 1998. BRDF models to predict spectral reflectance and emissivity in the thermal infrared. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 214–255.
- Sobrino, J. A., Jiménez-Muñoz, J. C., and Paolini, L., 2004. Land surface temperature retrieval from Landsat TM 5. *Remote Sensing of Environment*, **90**, 434–440.
- Tonooka, H., 2005. Accurate atmospheric correction of ASTER thermal infrared imagery using the water vapor scaling method. *IEEE Transactions on Geoscience and Remote Sensing*, **43**(12), 2778–2792.
- Tonooka, H., and Palluconi, F. D., 2005. Validation of ASTER/TIR standard atmospheric correction using water surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, **43**(12), 2769–2777.
- van de Griend, A. A., and Owe, M., 1993. On the relationship between thermal emissivity and the normalized difference vegetation index for natural surface. *International Journal of Remote Sensing*, **14**(6), 1119–1131.
- Wan, Z., 1999. MODIS land-surface temperature algorithm theoretical basis document (LST ATBD), Version 3.3. Contract NAS5-31370.
- Wan, Z., and Dozier, J., 1992. Effects of temperature-dependent molecular absorption coefficients on the thermal infrared remote sensing of the earth surface. In *Proceedings IGARSS’92*. pp. 1242–1245.
- Wan, Z., and Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(4), 892–905.
- Young, S. J., Johnson, R., and Hackwell, J. A., 2002. An in-scene method for atmospheric compensation of thermal hyperspectral data. *Journal of Geophysical Research*, **107**(24), 4774, doi:10.1029/2001JD001266.

Cross-references

Crop Stress
Cryosphere and Polar Region Observing System
Fields and Radiation
Global Programs, Operational Systems
Land Surface Emissivity
Ocean, Measurements and Applications
Optical/Infrared, Radiative Transfer
Sea Surface Temperature
Thermal Radiation Sensors (Emitted)
Volcanism
Water and Energy Cycles
Water Vapor

LAND SURFACE TOPOGRAPHY

G. Bryan Bailey

USGS Earth Resources Observation and Science Center,
Sioux Falls, SD, USA

Synonyms

Elevation; Landscape; Relief; Terrain

Definition

Topography of the Land Surface. The three-dimensional arrangement of physical attributes (such as shape, height, and depth) of a land surface in a place or region. Physical features that make up the topography of an area include mountains, valleys, plains, and bodies of water. Human-made features such as roads, railroads, and landfills are also often considered part of a region's topography (American Heritage Science Dictionary, 2005).

Importance of topographic information

The topography of the land surface is one of the most fundamental geophysical measurements of the Earth, and it is a dominant controlling factor in virtually all physical processes that occur on the land surface. Topography of the land surface also significantly controls processes within the overlying atmosphere, and it reflects the processes within the underlying lithosphere. Consequently, topographic information is important across the full spectrum of earth sciences.

Precipitation, runoff, soil moisture, incident sunlight, and temperature all vary with topography. Consequently, topography dominantly controls the local and regional distribution and character of vegetation. Erosion and sedimentation, and consequently soil formation and nutrient transport, also are strongly controlled by topography and are key factors in ecological studies. Topography strongly influences the location and magnitude of surface and subsurface water flow. Modeling of water supply and flood potential requires knowledge of the area's drainage extent, its slopes, and the pattern of the drainage network.

Particularly in rugged terrain, topography is commonly the dominant variable in remote sensing imagery. Topographic shading affects the radiance measured at every wavelength and is consequently the statistical principal component of many remotely sensed data sets. Meanwhile, atmospheric optical thickness varies inversely with topographic height, so that topography is an important factor in the atmospheric correction of remotely sensed data. Topographic data are imperative for the orthorectification of satellite imagery.

While topography controls many natural processes at and near the Earth's surface, many natural processes conversely control topography. Consequently, to various

degrees, topography records and reveals evidence of current and past natural processes. An obvious example is the development and occurrence of erosional and depositional fluvial landforms. Tectonic, volcanic, glacial, and gravitational processes also produce characteristic landforms that reveal past and ongoing change. Consequently, topographic information is an important tool in the study of such processes (Crippen, 2008).

Describing the topography of the land surface

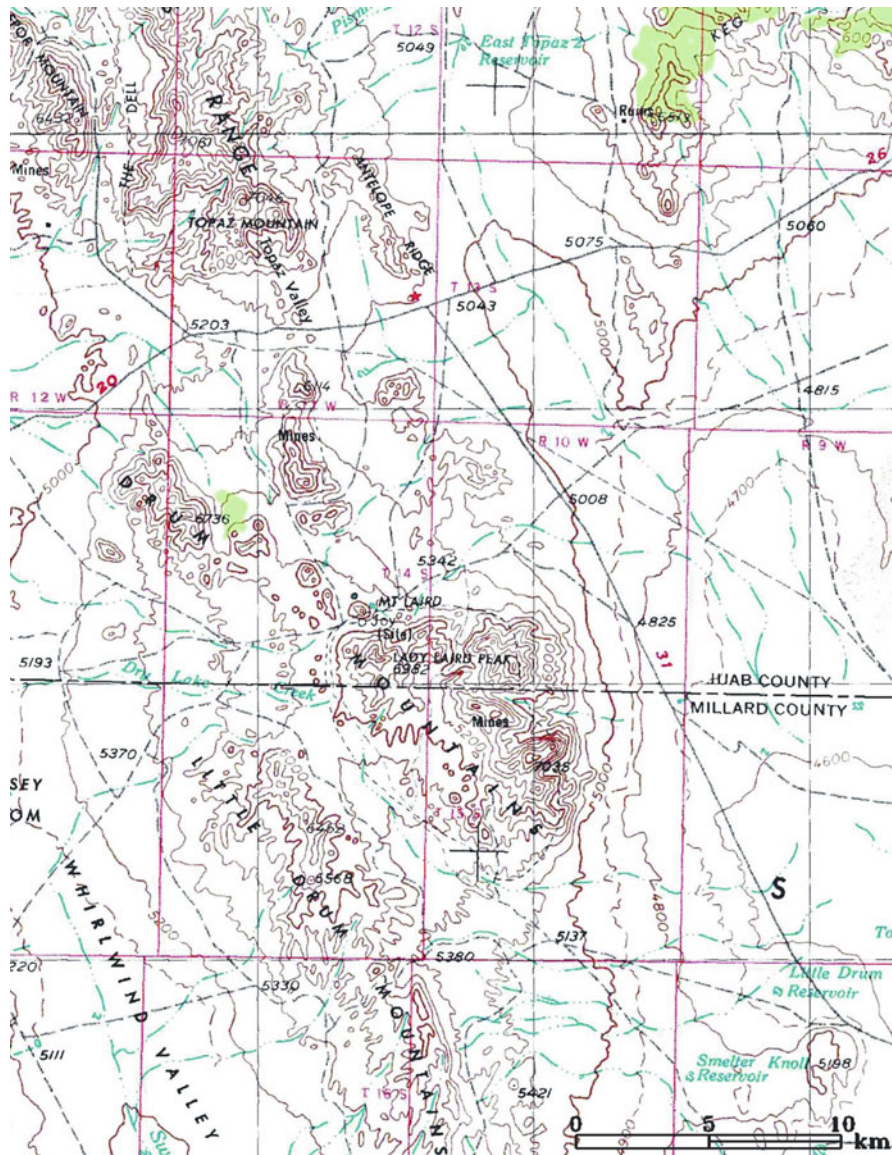
A *topographic map* (Figure 1) is a planimetric, or two-dimensional, representation of the three-dimensional configuration of the land surface where relief, or change in elevation, typically is represented by *contour lines*. A contour line is a line traced on the map such that all points on that line have the same elevation. That is, contour lines describe continuous points of equal elevation.

Until recent years, paper topographic maps were the most common tool used to describe the topography of the land surface. Until about 1940, most topographic maps were made by field crews who used alidades and plane tables to survey and map the topography of the landscape. World War II ushered in the age of *aerial photogrammetry* as the most common method for making topographic maps. This method, which uses overlapping and nadir-looking aerial photographs and a stereoplotter, revolutionized topographic mapping, resulting in greatly increased map coverage and enhanced map standardization (USGS, 1998).

Advances in computer technology brought about the latest great change in topographic mapping, the digital mapping revolution. Perhaps most notable was the replacement of the analog stereoplotter by the computer-assisted analytical stereoplotter. Not only has computer-assisted map production made it easier to make new paper maps and revise old ones, computer technology has accelerated demand for topographic data and other map information in *digital form* for use with the ever-growing number of computer-based mapping applications (McGlone, 2007).

A digital elevation model or DEM (Figure 2) is the generic term used most frequently to denote digital topographic data in all their various forms. The word "model" is applied because computers can use such data to model and automatically analyze the Earth's topography in three dimensions, thus avoiding much time-consuming human interpretation (Maune et al., 2007). Digital terrain model (DTM) and digital surface model (DSM) are two other common terms with similar meaning to DEM.

A *DEM* is a digital file consisting of terrain elevations for ground positions at regularly spaced horizontal intervals. The shorter those intervals are, the higher the spatial resolution of the DEM. The U.S. Geological Survey National Elevation Dataset (NED) typically has elevation data spaced at 30 m intervals, and it is thus said to have 30 m postings. DEMs are referenced to a vertical



Land Surface Topography, Figure 1 Part of USGS 1:250,000 topographic map of the Drum Mts., Utah, area.

datum, such as the WGS84/EGM96 geoid, and to a geographic coordinate system, such as Universal Transverse Mercator (UTM).

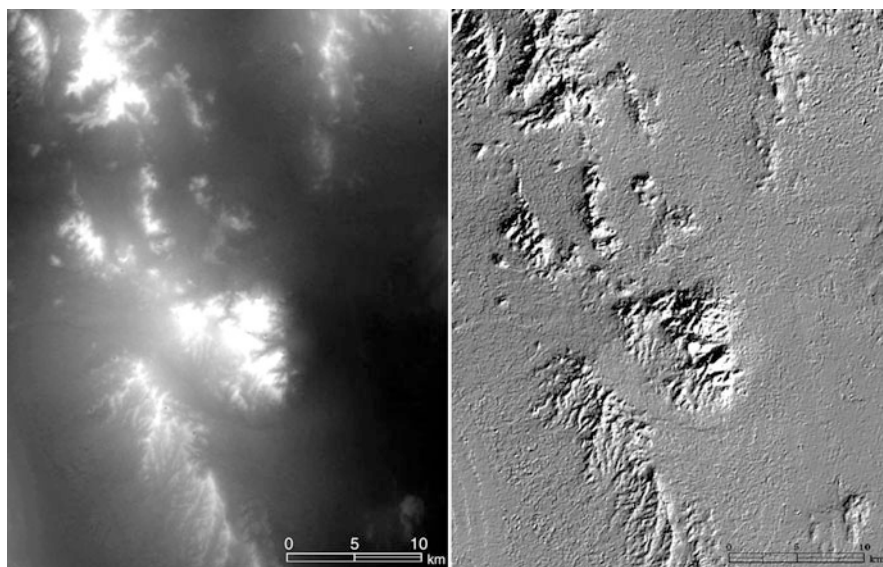
Determining land surface topography from remotely sensed digital data

Many DEMs have been, and continue to be, generated by digitizing topographic maps produced largely by aerial photogrammetric techniques. However, today most DEMs of the Earth's land surface are being generated, using a variety of automated processes, directly from digital data acquired by a rather large variety of airborne sensors and land surface-imaging satellite systems.

Most DEMs produced today from remotely sensed digital data are derived from one of three primary sources: optical imaging systems, interferometric synthetic aperture radar systems, and lidar systems (McGlone, 2007).

Optical imaging systems

Three types of optical imaging sensors are used for photogrammetric production of DEMs: airborne film cameras, airborne digital sensors, and digital sensors onboard satellites. *Film mapping cameras*, for decades the staple of aerial photogrammetric mapping, continue to be important sources of stereo images for DEM generation. Aerial film is scanned by high-resolution scanners to produce digital



Land Surface Topography, Figure 2 DEM intensity image (*left*) and DEM shaded relief image (*right*) of the Drum Mts., Utah.

images that can be processed by a softcopy stereoplotter or by one of the many available DEM generation software systems (McGlone, 2007).

Airborne and spaceborne digital sensors capable of collecting imagery useful in the generation of DEMs have certain similarities, but they also have important differences. Airborne systems are essentially digital mapping cameras capable of acquiring very high-spatial-resolution images, but typically with fewer spectral bands than spaceborne optical imaging systems. Stereo acquisition by airborne digital sensors typically is achieved by overlap of successive images acquired along the flight path of the aircraft, similar to film mapping cameras.

Optimal stereo coverage for DEM generation by satellite sensors is achieved along the orbital track of the satellite by using two (or more) sensors. One of the sensors is nadir looking, while the other points at some fixed angle along the orbital track fore and/or aft of the spacecraft. Some satellite systems acquire stereo imagery from adjacent orbits by pointing across track, and some are able to acquire limited stereo coverage along the same orbit using a single sensor that looks forward to image an area from one angle and then is pointed backward to image the same area from a different angle.

Generating DEMs from imagery acquired by airborne or spaceborne digital sensors is accomplished with the aid of a softcopy stereoplotter or one of the many available DEM generation software systems. The process may or may not employ the use of ground control points (GCPs), and it typically involves a sequence of steps that include selecting tie points in each of the stereo pair, co-registration of the stereo images, stereo correlation for parallax difference measurement (image matching), and calculation of elevation values. In the co-registered stereo

images, any positional differences between common points parallel to the direction of travel (parallax differences) are attributed to displacements caused by relief. Relative ground elevations are determined by measuring the parallax differences in the registered images, which then are converted to elevation (Lang and Welch, 1999).

Interferometric SAR systems

Synthetic aperture radars (SAR) illuminate the Earth's surface with microwave pulses, and they receive and record the return signals with respect to the magnitude and phase of those sine wave pulses (Bamler, 1997). While it is possible to generate accurate DEMs from stereo radar images using techniques similar to those described for optical imaging systems, DEMs are generated more commonly from *interferometric synthetic aperture radar* (InSAR).

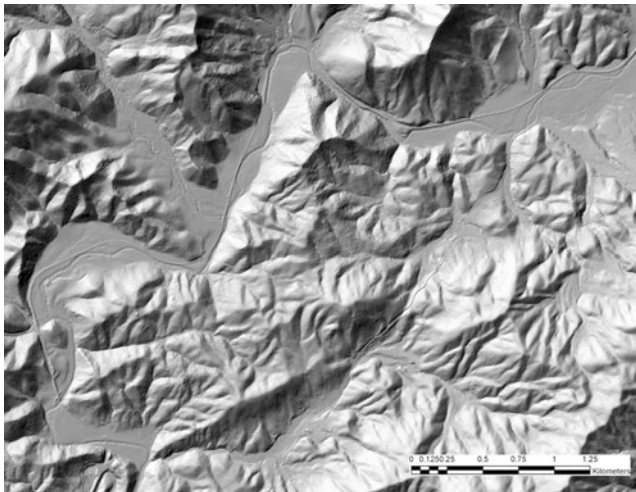
InSAR exploits the phase of SAR signals to measure stereo parallaxes to an accuracy of a fraction of a wavelength. The phase of the return radar wave depends on the distance to the ground, so it is possible to accurately determine land surface elevation on a pixel-by-pixel basis from the phase information. To generate a DEM, InSAR uses two SAR images of the same land surface area taken from slightly different positions and determines phase differences between them, producing an image called an *interferogram*. Further processing of the interferogram results in the generation of a DEM of the land surface imaged by the two sensors (Hensley et al., 2007).

Lidar systems

Lidar stands for Light Detection and Ranging, and like radar, it is an active remote sensing system. Lidars use

laser technology to measure distances to specific points by transmitting pulses or continuous waves of light, amplifying the light that is scattered back, and recording the precise time the transmitted pulse takes to travel to the target and back (Fowler et al., 2007). Most lidar systems used today to produce DEMs are airborne systems, and they employ a variety of different beam steering or scanning strategies. They also operate at a wide range of altitudes above the land surface, depending on the resolution requirements for the DEM to be produced. DEMs produced from lidar data typically have significantly greater spatial detail and better accuracy than DEMs produced from optical imaging systems or InSAR.

Lidar is a complex remote sensing technology, and the data processing required to convert raw lidar data to DEMs also is complex. Typically, such processing is done by the company or agency that collects that data, because the processing software has been developed specifically for the lidar system that collected the data (Fowler et al., 2007). The lidar product most commonly associated with topography is a DEM known as the *bare earth model*,



Land Surface Topography, Figure 3 Example of bare earth model DEM produced from lidar data collected over North Carolina.

which is a product from which the processing has removed virtually all returns not associated with the bare land surface (Figure 3).

Satellite systems that produce topographic data

This section briefly examines the system and data characteristics of selected Earth-observing *satellite systems* that acquire or have acquired data from which DEMs of the global land surface can be or have been generated. Airborne systems provide data from which highly accurate DEMs can be produced, but there is not one program or system that can provide such data for anywhere on the Earth's surface. The international remote sensing and earth science communities need access to quality DEMs for the entire global land surface, hence the emphasis here on satellite systems.

Radar satellite systems

Beginning in 1991, a number of polar-orbiting SAR satellite systems have been in operation, providing a continuous collective capability to acquire InSAR data from which DEMs can be generated for virtually any place on the global land surface. However, it was not until the 2010 launch of TanDEM-X, which operates in tandem formation with TerraSAR-X, that systematic efforts to produce a consistent global DEM from InSAR data were undertaken. Table 1 lists the SAR satellite systems that have contributed to DEM generation over the past two decades.

Shuttle radar topography mission (SRTM)

In 2000, the US National Aeronautics and Space Administration (NASA) and National Geospatial-Intelligence Agency (NGA) cooperated with the German Space Agency (DLR) to fly a Space Shuttle mission dedicated to acquiring digital topographic data for more than 80 % of the Earth's land surface. NASA flew a C-band SAR and DLR flew an X-band SAR, both of which were configured with one antenna in the bay of the space shuttle and the other at the end of a 60 m collapsible mast. The mission lasted 11 days, and NASA's C-band SAR collected complete InSAR coverage of the global landmass between 60 °N to 56 °S latitude. The X-band SAR was not designed for uninterrupted coverage.

Land Surface Topography, Table 1 International SAR systems from which DEMs can be produced from InSAR data

Country/agency	Satellite/sensor	Band	Launch date	Still active
European Space Agency (ESA)	ERS-1	C-band	1991	No
ESA	ERS-2	C-band	1995	No
Canadian Space Agency (CSA)	RADARSAT-1	C-band	1995	No
CSA	RADARSAT-2	C-band	2007	Yes
Japan (JAXA)	JERS-1	L-band	1992	No
ESA	Envisat ASAR	C-band	2002	No
Japan (JAXA)	ALOS PALSAR	L-band	2006	No
Germany (DLR)	TerraSAR-X	X-band	2007	Yes
Germany (DLR)	TanDEM-X	X-band	2010	Yes

The SRTM DEM data set is available from NASA and from the U.S. Geological Survey at no charge to users. Data are at 1 arc-second (30 m) for the United States and its territories and 3 arc-seconds (90 m) for all other covered areas. Vertical accuracy specification for the 1 arc-second data was 16 m (Rabus et al., 2003), but the data frequently have vertical accuracies better than 10 m (at 90 % confidence).

SPOT satellites

The French Système Pour l'Observation de la Terre (SPOT) satellites have been capable of acquiring cross-track stereo digital imagery from which DEMs can be generated since their initial launch in 1984. Spatial resolution of SPOT sensors has increased over that time, so the spatial details of DEMs that can be generated from those data likewise have increased. Currently, SPOT Image offers for sale DEMs produced from SPOT optical image data that cover most of the global land surface. The DEMs have 30 m postings and a vertical accuracy of less than 10 m where the slope of terrain is less than 20 %.

ASTER

The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) was built by the Japanese Ministry of Economy, Trade and Industry (METI), and it flies onboard NASA's Terra satellite. ASTER collects along-track stereo optical data with 15 m spatial resolution from which DEMs with 30 m postings are routinely produced as standard data products without the need for GCPs. The accuracy of ASTER DEMs varies some depending on terrain and other conditions, but they routinely have vertical accuracies better than 15 m (Fujisada et al., 2005). ASTER 60 km by 60 km DEMs are available for purchase to the general user from METI's Earth Resources Data Analysis Center and NASA's Land Processes Distributed Active Archive Center.

Since ASTER was launched in late 1999, more than two million scenes have been acquired of the global land surface. NASA and METI cooperated to produce a global DEM from these ASTER data. The *ASTER Global DEM*, which covers the Earth's landmasses from 83 °N to 83 °S latitude with nominal accuracies of 20 m vertical and 30 m horizontal at 95 % confidence, has 30 m postings. The ASTER Global DEM was contributed to the Global Earth Observing System of Systems (GEOSS) by NASA and METI, and thus it is available at no cost to users worldwide.

Cartosat and prism

Two optical imaging satellite systems that were designed to acquire data for generation of DEMs are the Indian Space Research Organisation's (ISRO) Cartosats and the Japanese Aerospace Exploration Agency's (JAXA) Prism sensor that flew onboard the Advanced Land Observation Satellite (ALOS). All acquire stereo image data along the orbital track. Cartosat-1 has a spatial resolution of 2.5 m

with a 30 km swath, and Cartosat-2 has a spatial resolution of less than 1 m with a 9.6 km swath. Prism has a 2.5 m spatial resolution and a 35 km swath. DEMs are not offered as standard products by either ISRO or JAXA, but the image data are available for purchase.

Summary and conclusions

The topography of the land surface is one of the most fundamental geophysical measurements of the Earth, and it is a dominant controlling factor in virtually all natural physical processes that occur on the land surface. The topographic map, which is a planimetric representation of the three-dimensional land, has been the most common tool used to describe the topography of the land surface until recently. Now, digital topographic data, in the form of a DEMs, are the tools of choice for many who wish to characterize the topography of the land surface.

DEMs are most frequently generated by automated computer techniques from digital data acquired by airborne and spaceborne sensors. Locally to regionally, airborne film and electro-optical systems and lidar systems provide users with high-quality and very accurate DEMs. For global studies, spaceborne optical systems capable of acquiring stereo imagery and InSAR systems offer the opportunity to produce DEMs with improving quality and accuracy worldwide. The SRTM DEM that cover 80 % of the global land surface and the ASTER Global DEM that covers virtually all of it are examples of two recent contributions by land remote sensing systems to better characterize the global land surface topography. Almost certainly, even greater advancements will be achieved in the next few years.

Bibliography

- Bamler, R., 1997. Digital terrain models from radar interferometry. In Fritsch, D., and Hobbie, D. (eds.), *Photogrammetric Week 1997*. Heidelberg: Wichmann Verlag, pp. 93–105.
- Crippen, R. E., 2008. Global topographic exploration and analysis with the SRTM and ASTER elevation models. In *Elevation Models for Geoscience*, Special Publication. London: Geological Society (in press).
- Fowler, R. A., Samberg, A., Flood, M. J., and Greaves, T. J., 2007. Topographic and terrestrial lidar. In Maune, D. F. (ed.), *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, 2nd edn. Bethesda: American Society of Photogrammetry and Remote Sensing, pp. 199–252.
- Fujisada, H., Bailey, G. B., Kelly, G., Hara, S., and Abrams, M., 2005. ASTER DEM performance. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 2715–2724.
- Hensley, S., Munjy, R., and Rosen, P., 2007. Interferometric synthetic aperture radar (IFSAR). In Maune, D. F. (ed.), *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, 2nd edn. Bethesda: American Society of Photogrammetry and Remote Sensing, pp. 141–198.
- Lang, H. R., and Welch, R., 1999. *ATBD-AST-08 Algorithm Theoretical Basis Document for ASTER Digital Elevation Models (Standard Product AST14)*. Washington, DC: National Aeronautics and Space Administration/Earth Observing System Program, p. 69.
- Maune, D. F., Kopp, S. M., Crawford, C. A., and Zervas, C. E., 2007. Introduction. In Maune, D. F. (ed.), *Digital Elevation Model Technologies and Applications: The DEM Users Manual*,

- 2nd edn. Bethesda: American Society of Photogrammetry and Remote Sensing, pp. 1–36.
- McGlone, J. C., 2007. Photogrammetry. In Maune, D. F. (ed.), *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, 2nd edn. Bethesda: American Society of Photogrammetry and Remote Sensing, pp. 119–140.
- Rabus, B., Eineder, M., Roth, A., and Bamler, R., 2003. The shuttle radar topography mission: a new class of digital elevation models acquired by spaceborne radar. *ISPRS Journal of Photogrammetry and Remote Sensing*, **57**, 241–262.
- The American Heritage® Science Dictionary Copyright © 2005 by Houghton Mifflin Company. Published by Houghton Mifflin Company.
- USGS, 1998. *Topographic Mapping*. Reston: U.S. Geological Survey. (Out of print; online version: <http://erg.usgs.gov/isb/pubs/booklets/topo/topo.html>).

Cross-references

Geodesy
 Geomorphology
 Land Surface Roughness
 Lidar Systems
 Radar, Synthetic Aperture
 Reflected Solar Radiation Sensors, Multiangle Imaging

LAND-ATMOSPHERE INTERACTIONS, EVAPOTRANSPIRATION

Joshua B. Fisher
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Synonyms

Evaporation; Water flux

Definition

Evapotranspiration (ET). The transfer of liquid water from open water and through plant transpiration to the atmosphere as water vapor.

Transpiration. The loss of water vapor through plant pores called stomata on leaves/needles or stems.

Basics of evapotranspiration

Evapotranspiration (ET) is the movement and transfer (i.e., flux) of water as a liquid at the Earth's surface to the atmosphere as a gas. ET is a combination of open water evaporation and plant transpiration. (Sublimation, which is the transition of solid water (i.e., ice, snow) to vapor due to low atmospheric pressure (i.e., high altitude), dry air, and high sunlight, is generally considered separate from ET.) Sources of open water evaporation could include oceans, seas, lakes, rivers, ponds, puddles, and water on objects such as plants, buildings, rocks, the soil surface (including movement of water vertically through the soil to the surface), or in the context of measuring devices such as a pan. Plants take up water from the soil through their roots, transferring that water through

their stems via conduits called xylem to their leaves, where it is used in the process of photosynthesis. The photosynthetic machinery in leaves (e.g., chlorophyll) takes in CO₂ from the atmosphere through stomatal pores and combines it with water and energy (i.e., light) to create sugars used to maintain and grow plant tissue and functions. While stomata are open, plants may lose water from their leaves to the atmosphere – this water loss is called transpiration. Plants regulate the opening and closing of their stomata to minimize water loss (closed), yet maximize CO₂ absorption (open).

Energy is required to break the strong bonds that hold water molecules together as a liquid – when those bonds break, the individual water molecules may enter the surrounding atmosphere as vapor. Energy may be in the form of heat, radiation, or pressure. Regardless of the availability of energy, water molecules may not be able to enter the atmosphere if the atmosphere is already saturated with moisture (humidity) or if there is no wind to facilitate the transfer of the molecules from the water source to the atmosphere. The wind itself may be differentially influenced by friction as it passes over smooth versus rough surfaces. Therefore, solar radiation (or, indirectly, air temperature), air humidity, and wind speed are the main climate influences on ET. The main vegetative controls include leaf and canopy structures, regulation of stomata, and rooting dynamics. Finally, soil characteristics control soil moisture retention of precipitation inputs. All of these potential controls vary in influence depending on the system in question, as well as the associated spatial and temporal scales of analysis (Fisher et al., 2011).

Remote sensing of ET

ET can be measured “remotely” with instruments attached to towers extending over vegetation using the eddy covariance technique (e.g., FLUXNET: Baldocchi et al., 2001). These same instruments may be attached to airplanes for regional measurements. However, ET cannot be measured directly from satellite remote sensing, so it must be inferred from a model or the residual of other measurements. There are three orders of complexity in space-based estimation of ET:

- Simple: Empirical, semiempirical
- Intermediate: Water balance, energy balance
- Complex: Land surface/Earth system models

Empirical, semiempirical approaches

One of the simplest approaches to estimating ET is to take another closely related variable that is measureable and convert that to ET using a statistical relationship. The statistical relationship (e.g., linear regression) may be developed from studies where both the other variable and ET were measured and then used to extrapolate beyond the site. One commonly used variable is the Normalized Difference Vegetation Index (NDVI), as well as related “greenness” indices, constructed from

measurements primarily in the red and near-infrared (NIR) wavelengths, and which is indicative of plant productivity. Where there is plenty of water and energy, there will be both high NDVI and ET; where there is no water and energy, one would not expect much NDVI and ET. However, this relationship may fall apart, for example, under deforestation or nutrient limitation (high ET, low NDVI), or diurnal/seasonal water stress (low ET, high NDVI). NDVI may be obtained from satellite instruments such as the Advanced Very High Resolution Radiometer (AVHRR), the Moderate resolution Imaging Spectroradiometer (MODIS), or the Visible Infrared Imager Radiometer Suite (VIIRS).

Another commonly used variable is Land Surface Temperature (LST), constructed from thermal-infrared (TIR) measurements. A given surface may be cooled (lower LST) when evaporating and hotter when there is less ET. However, other forces may change the temperature of the surface, including advecting warm/cool/dry/moist air. LST may be obtained from satellite instruments such as MODIS, the Atmospheric Infrared Sounder (AIRS), or Landsat. LST may be combined with NDVI for a somewhat more complex empirical approach. One of the leading empirical approaches comes from the MPI-BGC product, which is constructed from a machine learning technique and model tree ensemble that developed statistical relationships between measured ET and globally available ancillary data at over 250 FLUXNET sites (Jung et al., 2009). Finally, many agriculturalists use semiempirical algorithms to estimate ET, using physics-based equations for potential ET (PET), then converting or downscaling PET to actual ET (AET) using an empirical scalar multiplier, called a crop coefficient, developed for their specific crop and location.

Water balance

ET may be calculated as the residual of known measurements in the water balance equation:

$$P = dS + Q + ET \quad (1)$$

where P is precipitation (rainfall and snow); dS is the change in stored standing water (e.g., lakes, ponds, or in/on plants), soil moisture, and groundwater; and Q is runoff. From a remote sensing standpoint, rainfall is measured from a variety of satellites including the Tropical Rainfall Measuring Mission (TRMM) (the Global Precipitation Mission (GPM) is currently in development as the next major multi-satellite precipitation-measuring mission) and snow from MODIS. dS is measureable at large spatial scales from the Gravity Recovery And Climate Experiment (GRACE). Q is not yet measureable from space, (The proposed Surface Water Ocean Topography (SWOT) mission currently in development would measure river discharge from space.) but is readily obtained from river discharge measurements, though many rivers are sparsely instrumented, for

example, in developing nations. Equation 1 may be rearranged to solve for ET given the known measurements of the three other variables in the equation.

Energy balance

ET may also be considered an energy (water fluxes such as precipitation and ET are usually given in units of depth per time (i.e., $\text{mm} \cdot \text{day}^{-1}$); the units are consistent when they are in volume per area per time (i.e., $\text{m}^3 \cdot \text{ha}^{-1} \cdot \text{day}^{-1}$). 1 m^3 is equal to 1,000 l. Water can also be expressed in units of mass – 1 kg of water is equal to 1 mm of water spread over 1 m^2 . ET , like R_n , can be expressed in units of energy too. Because it requires 2.45 MJ to vaporize 1 kg of water (at 20°C), 1 kg of water is therefore equivalent to 2.45 MJ; 1 mm of water is thus equal to $2.45 \text{ MJ} \cdot \text{m}^{-2}$) variable, called the latent heat of evaporation (LE), as it requires a certain amount of energy to convert a given quantity of liquid water to gas. Energy coming from the sun less any radiation that gets reflected back to the atmosphere – or net radiation (R_n) – is energy available to drive ET. Any R_n that does not drive ET either gets converted to sensible heat (H) or stored in the soil or other objects (G):

$$R_n = ET + H + G \quad (2)$$

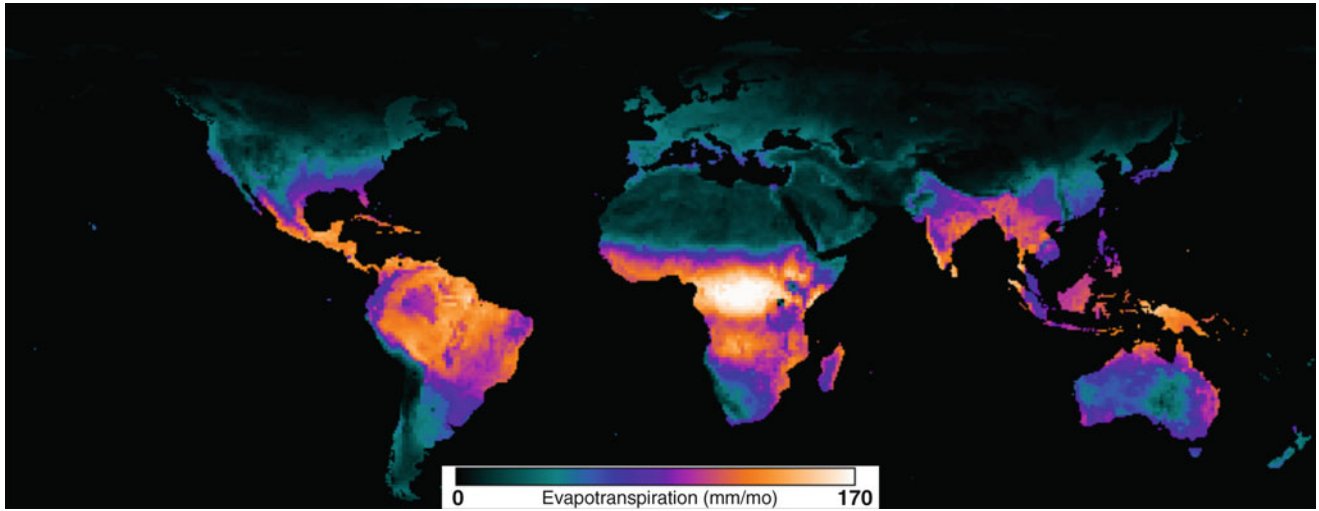
A few space-based R_n are available, including those from the Surface Radiation Budget (SRB), the Clouds and Earth's Radiant Energy System (CERES), the International Satellite Cloud Climatology Project (ISCCP), and MODIS. H and, to a lesser extent, G are not remotely measureable and are the focus of models such as the Surface Energy Balance System (SEBS), the Atmosphere-Land Exchange Inverse (ALEXI), the Surface Energy Balance Algorithm for Land (SEBAL), and Mapping EvapoTranspiration at high Resolution with Internalized Calibration (METRIC), all of which rely particularly on remotely sensed LST (Li et al., 2009).

Direct approaches

ET may also be calculated “directly” from the physics that control ET, as outlined earlier in the “Basics of Evapotranspiration” subsection. The most widely used equation for determining ET comes in the form of the Penman-Monteith equation:

$$ET = \frac{\Delta R_n + \frac{c_p \rho VPD}{r_a}}{\Delta + \gamma + \gamma \left(\frac{r_s}{r_a} \right)} \quad (3)$$

where Δ is the slope of the saturation-to-vapor pressure curve, c_p is the specific heat of water, ρ is air density, VPD is vapor pressure deficit, r_a is aerodynamic resistance, γ is the psychrometric constant, and r_s is surface resistance. Equation 3 forms the foundation of the algorithm for the official MODIS ET product (MOD16) (Mu et al., 2011), which relies on MODIS-based leaf area index (LAI), fraction of absorbed photosynthetically active radiation (fAPAR), land cover, and a general



Land-Atmosphere Interactions, Evapotranspiration, Figure 1 Mean monthly ET for 2004 from the PT-JPL product.

biome-specific lookup table to parameterize the resistances. R_n , VPD , and air temperature (T_a ; i.e., included in Δ) in MOD16 are derived from the NASA/GMAO Modern Era Retrospective Analysis (MERRA).

The PT-JPL product (Figure 1: Fisher et al., 2008) is based on the PET formulation of the Priestley-Taylor equation, which is a reduced version of the Penman-Monteith equation, eliminating the need to parameterize the stomatal and aerodynamic resistances, leaving only equilibrium evaporation multiplied by a constant (1.26) called the α coefficient:

$$PET = \alpha \frac{\Delta}{\Delta + \gamma} R_n \quad (4)$$

PET is reduced to AET using ecophysiological constraint functions (f -functions, unitless multipliers, 0–1) based on atmospheric moisture (VPD and relative humidity, RH) and vegetation indices ($NDVI$ and $SAVI$):

$$ET = ET_s + ET_c + ET_i \quad (5)$$

$$ET_c = (1 - f_{wet}) f_g f_T f_M \alpha \frac{\Delta}{\Delta + \gamma} R_{nc} \quad (6)$$

$$ET_s = (f_{wet} + f_{SM})(1 - f_{wet}) \alpha \frac{\Delta}{\Delta + \gamma} (R_{nc} - G) \quad (7)$$

$$ET_i = f_{wet} \alpha \frac{\Delta}{\Delta + \gamma} R_{nc} \quad (8)$$

where ET_s , ET_c , and ET_i are evaporation from the soil, canopy, and intercepted water, respectively, each calculated explicitly. f_{wet} is relative surface wetness (RH^4), f_g is green canopy fraction (f_{APAR}/f_{IPAR}),

f_T is a plant temperature constraint ($\exp(-((T_{max} - T_{opt})/T_{opt})^2)$), f_M is a plant moisture constraint ($f_{APAR}/f_{APARmax}$) and f_{SM} is a soil moisture constraint, (RH^{VPD}). f_{APAR} is absorbed photosynthetically active radiation (PAR), f_{IPAR} is intercepted PAR, T_{max} is maximum air temperature, T_{opt} is T_{max} at $\max(R_n T_{max} SAVI/VPD)$, and G is the soil heat flux.

Land surface models/Earth system models

The most complex approach to estimating ET is through full Land Surface Models (LSMs) or Earth System Models (ESMs). These models are typically driven by meteorological data and aim to simulate all of the relevant biogeochemical processes and states governing the exchange of energy, water, and carbon throughout the entire land surface or complete Earth system, including ocean and atmosphere. Some of these models assimilate any relevant observation from both space and in situ to constrain the complexity of linkages and feedbacks. While the estimate of ET from LSMs and ESMs is subject to potentially greater uncertainty relative to the previously described approaches due to increased complexity and degrees of freedom, LSMs and ESMs allow more realistic feedbacks to and from ET given changes in the Earth system or climate (Mueller et al., 2011).

Summary

Remote sensing of ET is currently a high-level research and science priority, especially as ET is central to connecting the water, energy, and carbon cycles; a modulator of regional rainfall; a significant factor in flood and drought processes and models; the primary climatic predictor of biodiversity; and critical for the agricultural industry. In situ measurement of ET requires

cost-constraining equipment; as such, major international efforts, such as the Global Energy and Water Cycle Experiment (GEWEX), have focused on determination of ET from existing remote sensing assets (Jiménez et al., 2011; Vinukollu et al., 2011). The techniques described here provide an overview of how the scientific community estimates ET from remote sensing.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Baldocchi, D., et al., 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, **82**, 2415–2434.
- Fisher, J. B., Tu, K., and Baldocchi, D. D., 2008. Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sensing of Environment*, **112**, 901–919.
- Fisher, J. B., Whittaker, R. H., and Malhi, Y., 2011. ET come home: a critical evaluation of the use of evapotranspiration in geographical ecology. *Global Ecology and Biogeography*, **20**, 1–18.
- Jiménez, C., et al., 2011. Global inter-comparison of 12 land surface heat flux estimates. *Journal of Geophysical Research*, **116**, 1–27, doi:10.1029/2010JD014545.
- Jung, M., Reichstein, M., and Bondeau, A., 2009. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, **6**, 2001–2013.
- Li, Z.-L., et al., 2009. A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. *Sensors*, **9**, 3801–3853.
- Mu, Q., Zhao, M., and Running, S. W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, **111**, 519–536.
- Mueller, B., et al., 2011. Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophysical Research Letters*, **38**, 1–7, doi:10.1029/2010GL046230.
- Vinukollu, R. K., Wood, E. F., Ferguson, C. R., and Fisher, J. B., 2011. Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: evaluation of three process-based approaches. *Remote Sensing of Environment*, **115**, 801–823.

LANDSLIDES

Vernon H. Singhroy
Applications Development Section, Natural Resources
Canada, Canada Centre for Remote Sensing, Ottawa,
ON, Canada

Synonyms

Debris flow; Mass movement; Mudslide; Rock avalanche

Definition

Landslide is used to describe the downslope movement of soil and rock under the effects of gravity. In some cases, other terms such as mass movements and slope failure are used interchangeably with landslides. The most common triggers of landslides are earthquakes, heavy rains, thawing of frozen ground, river and coastal erosion, and frequent infrastructure and building development. Many types of landslides are usually associated with specific mechanics of slope failure and the properties and characteristics of failure type. Figure 1 shows a simple illustration of a rotational landslide which illustrates the commonly used labels for the parts of a landslide.

Introduction

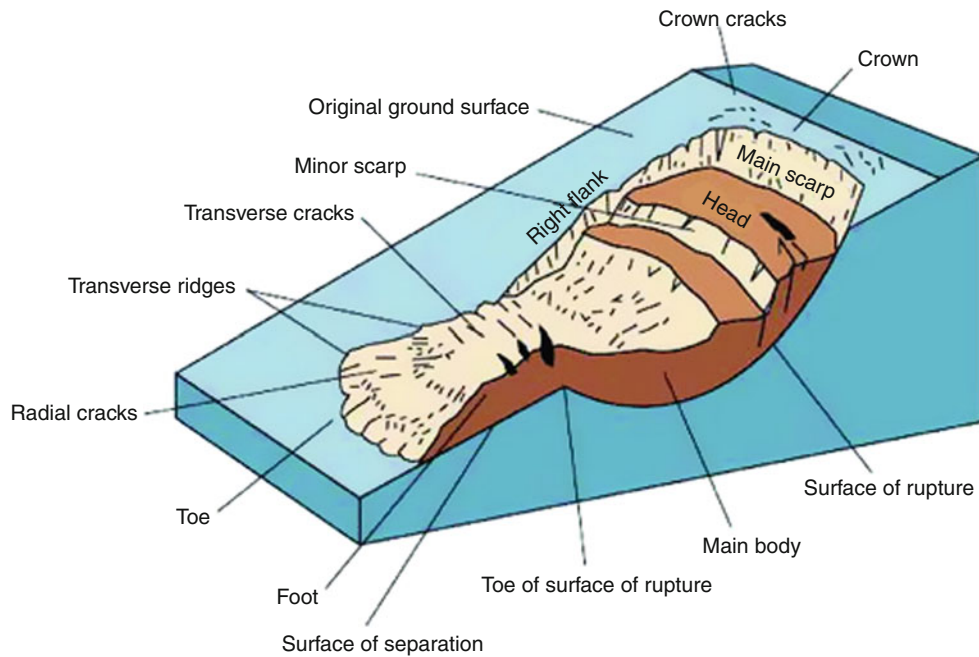
Landslides are among one of the serious geological hazards which threaten and influence the socioeconomic conditions of many countries (Schuster, 1996). They are the manifestation of slope instability. An example of their destructive nature is shown in Figure 2. The La Conchita landslide in California killed ten persons. Geologists and engineers have long tried to identify the conditions of slope failure and mitigate their risk to infrastructure and populated areas.

There are various techniques used to map and evaluate landslides. Large-scale, stereo aerial photographs are one of these tools that have been extensively used in landslide investigations. Because of their three-dimensional capability, they provide essential geologic and geomorphic information necessary for landslide inventory mapping. Recently, there has been increasing uses of high-resolution satellite images (1–5 m) for various landslide investigations. The recent advances in radar interferometry are providing valuable insights in monitoring slow-moving landslides. The following section briefly discusses the uses of these techniques in landslide investigations.

Landslide mapping

Stereo aerial photographs are used extensively to produce landslide inventory maps. They allow the identification of geomorphic, geologic, and related land use features related to landslides (Mollard and Janes, 1993). Geological and geomorphologic units related to landslide inventories can be interpreted on the basis of morphological, textural, and structural characteristics using stereo aerial photos and high-resolution satellite images. Landslide inventory maps are usually published at various scales, such as national (1; 1,000,000), regional (1:100,000), medium (1:25,000–50,000), and large scales (>1:15,000).

For instance, detail inventory requires detail aerial photos and high-resolution satellite images to assist the interpreter to make conclusions on types and causes of the landslide. Recently, the high-resolution satellite Google images provided a cheap and valuable source of



Landslides, Figure 1 Parts of a landslide (Modified from Varnes (1978), by Highland and Bobrowsky (2008)).

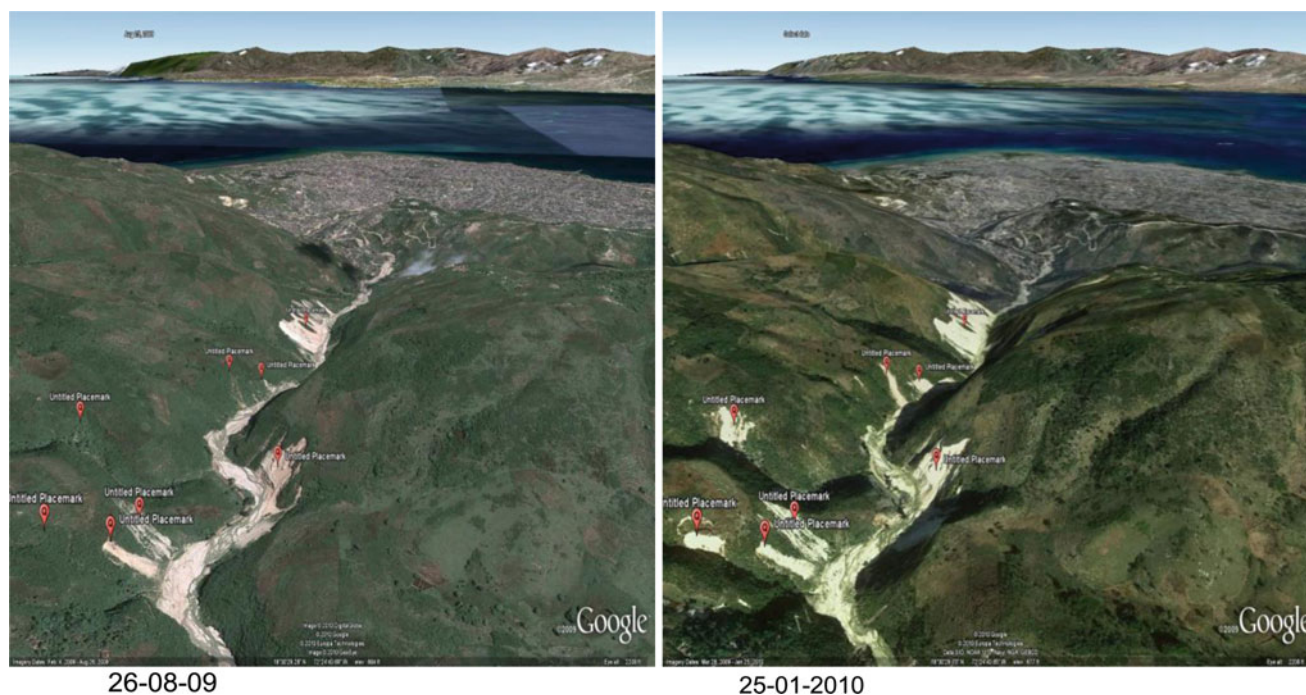


Landslides, Figure 2 This landslide occurred at La Conchita, California, USA, in 2005. Ten people were killed (USGS Photo).

locating landslides before and after the devastating earthquake in Haiti in January 2010 (Figure 3).

Other high-resolution optical systems (1–5 m) such as IKONOS, Quickbird, and IRS images and the stereo capability of SPOT 5 are useful for landslide recognition

and related land use mapping. Whenever possible, the highest-resolution images should be used to identify and interpret the geomorphic and associated features shown in Figure 1. Large landslides are easily recognized from medium resolution 30 m Landsat TM images.



Landslides, Figure 3 Landslides near Port-au-Prince, before and after Haiti earthquake.

Recent research has shown that high-resolution stereo radar and optical images, combined with topographic and geological information, have assisted in mapping the geomorphic characteristics of deep-seated landslides needed to produce of landslide inventory maps (Singhroy et al., 1998; Singhroy, 2005). The multi-incidence angle, stereo, and high-resolution capabilities of the various radar satellites are particularly useful in providing terrain and geomorphic information needed to produce landslide inventory maps.

Currently, damage assessment related to landslides and other disasters in support of relief efforts uses aerial photography, videos, high-resolution satellite images, and ground checks.

Landslide monitoring using InSAR

Landslides usually resulted in extreme economic and societal costs, despite our increased understanding of the mechanisms of failure and large ground deformation. Current state of the art in real-time monitoring of active slopes developed for early warning of landslides is very expensive. Satellite radar interferometry is used increasingly to complement real-time monitoring such as GPS and in situ field measurements (Singhroy, 2008).

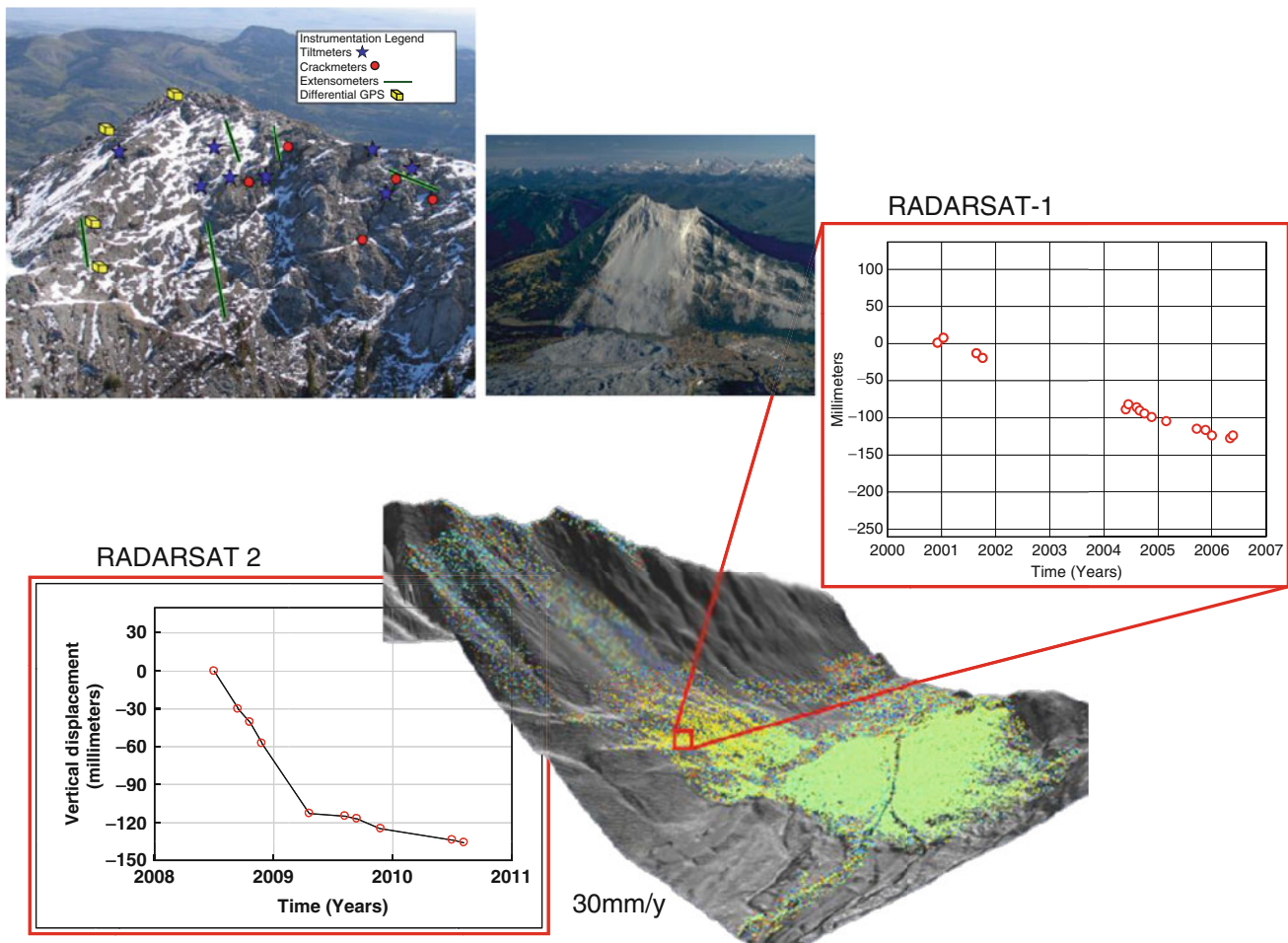
Interferometric Synthetic Aperture Radar (InSAR) techniques are being used to measure small millimeter displacement on slow-moving landslides. An interferometric image represents the phase differences between the backscatter signals in two SAR images obtained from similar positions in space. In case of spaceborne SAR, the images are acquired

from repeat-pass orbits. The phase differences between two repeat-pass images result from topography and from changes in the line-of-sight distance (range) to the radar due to displacement of the surface or change in the atmospheric propagation path length. For a nonmoving target, the phase differences can be converted into a digital elevation map if very precise satellite orbit data are available.

Typical scales for SAR interferometry application to landslide movements are millimeters to centimeters per orbit cycle of the radar satellite. This orbit cycle can range from 44 days for ALOS and 10 days for TerraSAR-X. Constellation missions such as Cosmo-SkyMed, Sentinel, and RADARSAT constellation mission are reducing the orbits to 1–4 days.

It is clear that InSAR techniques can be used to monitor landslide motion under specific conditions, provided coherence is maintained over the respective orbit cycle. Using data pairs with short perpendicular baselines, short time intervals between acquisitions, and correcting for the effect of topography and atmospheric effects, reliable measurements of surface displacement can be achieved. The InSAR deformation maps provide linear motion at the line of site. Although this is very useful information, landslide motion is very complex with nonlinear vectors. Therefore, InSAR techniques do not provide the complete 3d motion

Vegetation decorrelates the radar signals. Therefore, stable coherent targets such as installed corner reflectors or man-made constructions such as houses, roads, and bridges are used to calculate the landslide motion. The uses of installed field corner reflectors are increasing on remote vegetated sites. Acquiring about 30 InSAR images on coherent targets over long periods are analyzed by a Permanent



Landslides, Figure 4 Frank Slide, Canada: monitoring landslide motion from RADARSAT InSAR (2000–2010) (Modified from Mei et al. 2007).

Scatterer Technique PSInSAR™ developed by Ferretti et al. (2001). With the Permanent Scatterer Technique, the movement of small objects (down to about 1 m²) can be monitored. This method has been applied to map subsidence and slow-moving landslides and many parts of the world.

The *Frank rock avalanche* is provided to demonstrate the capability of InSAR to monitor gradual motion on large rock avalanche in the Canadian Rockies. The Frank Slide, a 30 × 106 m³ rock avalanche of Paleozoic limestone, occurred in April 1903 on the east face of Turtle Mountain in southern Alberta, Canada. Seventy fatalities were recorded. This slide is still active. Several factors contribute to this rock avalanche. These include the geological structure of the mountain, subsidence from coal mining at the toe of the mountain, blast-induced seismicity, above-average precipitation in years prior to the slide, and freeze-thaw cycles (Cruden and Hungr, 1996). GPS stations and several in situ monitors are installed to monitor post-slide activity at specific

locations (Figure 4). Current InSAR monitoring is complementing the in situ measurements. The fact that the rock covering the rock avalanche is bare and dry leads to the high coherence and identification of more than 95 % of the coherent target monitoring targets for the Frank Slide area. Due to their great density and excellent coverage, the coherent target measurements of this area are a reliable reflection of current deformation pattern. The most recent InSAR results (Figure 4) have shown that during a period from April 2004 to October 2006, the foot of the eastern slope of Turtle Mountain, the ground surface above the coal mine, was found to subside at an average rate of about 3.1 mm per year supporting the speculation that underground coal mining triggered the Frank landslide (Mei et al., 2007).

The above examples show that satellite images are providing reliable complementary techniques to landslide mapping and monitoring, and therefore, its uses are increasing in landslide investigation and mitigation.

Bibliography

- Cruden, D. M., and Hungr, O., 1996, 1986. The debris of Frank Slide and theories of rockslide-avalanche mobility. *Canadian Journal of Earth Sciences*, **23**, 425–432.
- Ferretti, A., Prati, C., and Rocca, F., 2001. Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 8–20.
- Highland, L. M., and Bobrowsky, P., 2008. *The Landslide Handbook – A Guide to Understanding Landslides*. Reston: U.S. Geological Survey Circular, Vol. 1325, p. 129.
- Mei, S., Poncos, V., and Froese, C., 2007. *InSAR Mapping of Millimetre-scale Ground Deformation over Frank Slide, Turtle Mountain, Alberta*. Alberta Energy and Utilities Board, EUB/AGS Earth Science Report 2007, 1–62 pp.
- Mollard, J. D., and Janes, J. R., 1993. *Airphoto Interpretation of the Canadian Landscape*. Ottawa: Energy, Mines and Resources, Canada, p. 415p.
- Schuster, R. L., 1996. Socio- economic significance of landslides. In Turner, A. K., and Schuster, R. L. (eds.), *Landslides: Investigation and Mitigation*. Washington, DC: National Academy Press. Report 247, Transportation Research Board, NRC, pp. 12–35.
- Singhroy, V., 2005. Remote sensing for landslide assessment: chapter 16. In Glade, T., Anderson, M., and Crozier, M. J. (eds.), *Landslides Hazard and Risk*. Chichester/Hoboken: Wiley, pp. 469–549.
- Singhroy, V., 2008. Satellite remote sensing applications for landslide detection and monitoring (chap. 7). In Sassa, K., and Canuti, P. (eds.), *Landslide Disaster Risk Reduction*. Berlin: Springer, pp. 143–158.
- Singhroy, V., Mattar, K. E., and Gray, A. L., 1998. Landslide characteristics in Canada using interferometric SAR and combined SAR and TM images. *Advances in Space Research*, **3**, 465–476.
- Varnes, D. J., 1978. Slope movement, types and processes. In Schuster, R. L., and Krizek, R. J. (eds.), *Landslides-Analysis and Control*. Washington, DC: National Research Council. Transportation Research Board Special Report 176, pp. 11–23.

LAW OF REMOTE SENSING

Joanne Irene Gabrynowicz

National Center for Remote Sensing, Air, and Space Law,
The University of Mississippi School of Law, Mississippi,
MS, USA

Overview

All of the international space law began with the Declaration of Legal Principles Governing the Activities of States in the Exploration and Use of Outer Space (Declaration), adopted in 1962 by the United Nations General Assembly. National space laws, like that of the United States, were influenced by international space law and developed in tandem.

The Declaration is the foundation for the Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (Outer Space Treaty). The Outer Space Treaty entered into force at the height of the Cold War on October 10, 1967. It contains fundamental space

law principles that are directly applicable to remote sensing activities, such as all nations have the nonexclusive right to use space. In less than a decade, four more treaties followed, some of which also have legal principles applicable to remote sensing.

Legislative history in the United Nations

In 1970, Prof. A.A. Cocca of Argentina first introduced remote sensing as a specific legal topic in a paper to the Legal Subcommittee of the U.N. Committee on the Peaceful Uses of Outer Space (UNCOPUOS). In 1971, a Working Group was formed in the Legal Subcommittee to consider the paper, and in 1973, the Scientific and Technical Subcommittee of UNCOPUOS issued its first report containing a section on remote sensing. In 1974, the General Assembly adopted a resolution recommending that the LSC should consider the question of the legal implications of remote sensing of the Earth from space.

In 1975–1976, the first discussions concerning the legal implications of remote sensing began. Initially, the participating Nation-States organized themselves into three groups, or blocs: the Soviet Union/East Europe; the Group of 77 (G-77), consisting of developing nations in Africa, Asia, and Latin America; and the Western Group. The primary issues were whether or not sensing States had to obtain the consent of a sensed State prior to acquiring data from space or prior to distributing data to a third party. The Soviet Union and France submitted a proposal to require sensing States to obtain the prior consent of sensed States before data could be made available to other entities. The G-77 initially opposed both remote sensing itself and the distribution of data. The United States, then the only sensing state, advocated a free flow of data and therefore opposed prior constraints.

In 1975, the General Assembly recommended that the Legal Subcommittee continue consideration of remote sensing from space as a high priority. It specifically pointed to the use of remote sensing regarding the Earth's natural resources and environment. It also recommended drafting principles regarding points on which States agreed. The General Assembly noted that the Scientific and Technical Subcommittee had examined operational and experimental questions and now recommended that studies be conducted on organizational and financial matters. It also endorsed an international remote sensing training center for personnel from developing nations. In 1976, this work slowed down because the Soviet Union attempted to link some remote sensing issues to the developing Moon Treaty. By late 1976, progress was made on formulating some draft principles. The General Assembly noted in a resolution that the Legal Subcommittee formulated five draft principles and identified three new common elements identified by States.

Main issues

From 1977 to 1979, the Working Group focused on three main issues: whether or not “should” or “shall” ought to

be used in the principles, providing consultation and dispute resolution procedures, and the Soviet Union's proposal to limit image gathering to 50 m spatial resolution. Cold War politics drove the legal debate. A specific attempt to codify the 50 m limit was made on May 19, 1978, when Cuba, Czechoslovakia, the German Democratic Republic, Hungary, Mongolia, Poland, Romania, and the Union of Soviet Socialist Republics signed the Convention on Transfer and Use of Data of Remote Sensing of the Earth from Outer Space in Moscow. It provided that if a Contracting Party was in possession of data with resolution higher than 50 m, it was forbidden from making the data available to anyone without the explicit consent of the sensed State. This did not significantly influence the Legal Subcommittee negotiations. In a 1979 resolution, the General Assembly again recommended that the Legal Subcommittee continue work on the draft principles on a priority basis.

The evolution and use of the US *Landsat* system catalyzed addressing many general and specific remote sensing issues in legal terms. In 1982 and 1983, as a cost recovery method, the United States raised the access fee for a *Landsat* ground station from \$200,000 (US) to \$600,000 (US), and the cost of computer-compatible tapes increased. The United States announced its intention to commercialize the *Landsat* system. Identifying principles regarding [data access](#), among others, became more pressing. In 1981, 1982, 1983, and 1984, the General Assembly adopted resolutions that noted each year's progress toward developing principles and continued to urge the Legal Subcommittee to develop the legal implications of remote sensing on a priority basis.

In 1984, the US Congress passed the Land Remote Sensing Commercialization Act (Commercialization Act). It adopted the nondiscriminatory access policy forged in the Legal Subcommittee and provided for a three-phased process to establish commercial remote sensing. The first phase was to award a contract for existing *Landsat* operations to a private sector operator. The second phase was to be a transition period in which both the government and the private sector would operate satellites, with government activities phasing out. The third phase was to be a fully private, commercial environment. The US contract award process had begun and the negotiators in the United Nations took note. Conversely, the United States needed to have the legitimacy of commercial remote sensing activities accepted. France was preparing to launch its first remote sensing satellite, *SPOT-1*, which it did in 1986. In 1984, France proposed alternate language for the draft data access principle, and negotiations on the remote sensing principles were revitalized.

In 1985, significant portions of the draft principles were still not agreed upon. According to at least one report, the Soviet Union/Eastern bloc did not participate in discussions of Article XII, the data access principle, out of concern that there would be no acceptable solutions for the G-77 nations. Some nations, including Mexico and

Vietnam, believed there was insufficient time to consider the draft text. The General Assembly again adopted a resolution endorsing the Legal Subcommittee's work and added that it should finalize the draft set of principles. The Chair of the Working Group was Austria, and it offered an alternate text based on consultations regarding the French proposal. Discussions on the Austrian text were held. No changes were made in 1986. On December 3, 1986, The Principles Relating to Remote Sensing of the Earth from Outer Space (Principles) were adopted by the General Assembly.

The principles relating to remote sensing of the Earth from outer space

The Principles embody the view that Outer Space is a resource for all humanity and should be used for the general benefit of all nations. They encourage international cooperation and address access and distribution of data and information generated by national civilian remote sensing systems. Primary data are the raw data delivered in the form of electromagnetic signals, photographic film, magnetic tape, or any other means. Processed data are the products resulting from processing primary data, and analyzed information means information resulting from interpreting processed data. Remote sensing activities include operations, data collection, storage, processing, interpretation, and dissemination.

The Principles set a standard of international cooperation among sensing and sensed States while attempting to achieve a balance between their rights and interests. Needs of developing nations are given special regard. The Principles specifically promote protection of the Earth's environment and of humanity from natural disasters. States that possess remotely sensed information useful for averting harmful phenomena are required to disclose the information to concerned States. If the potential harm threatens people, the obligation to disclose requires promptness and extends to processed data and analyzed information.

The rights and responsibilities of sensed and sensing States are particularly addressed in Articles IV and XII. Article IV sets a legal standard for behavior among sensed and sensing States, and Article XII is a data dissemination statute. Together, they provide a fluid legal regime that obliges sensing States to avoid harm to sensed States and to provide them with access to primary data and processed data concerning their own territory on a nondiscriminatory basis. This was the compromise between terrestrial sovereignty and the freedom to use space. The legitimacy of space-based remote sensing was accepted by ensuring that a sensed State would have access to the imagery of its territory. Analyzed information available to sensing States is also to be available to the sensed States on the same basis and terms. In turn, sensed States are to meet reasonable cost terms and do not have access to analyzed information legally unavailable to the sensed States, for example, proprietary information.

The Principles were contained in the first major resolution to emerge from COPUOS in more than a decade and provide a foundation for the continued evolution of international remote sensing law. The question of whether or not the remote sensing principles ought to become a treaty continues to be raised in the COPUOS Legal Subcommittee.

US law: The oldest national remote sensing law

On October 28, 1992, the US Congress passed the Land Remote Sensing Policy Act of 1992 (Policy Act). It replaced the 1984 Commercialization Act. Congress adopted the nondiscriminatory access policy for second time. The Policy Act's focus is long-term remote sensing policy and its numerous facets. Specific matters addressed include program management, *Landsat 7* procurement, *Landsat 4* through *7* data policy, transfer of *Landsat 6* program responsibilities, regulatory authority and administration of public and private remote sensing systems, federal research and development, advanced technology demonstration, *Landsat 7* successor systems, data availability and archiving, and the continued prohibition of weather satellite commercialization. The legislation features a focus on the value of remote sensing in conducting global change research and other public sector applications, a recasting of remote sensing activities, and provisions for the future evolution of remote sensing policy. In 2008, efforts were made to replace the 1992 law with a new statute titled, the *National Land Imaging Program*. This bill was intended to embody the US new national remote sensing policy to implement a long-term operational land imaging program. It was not made into law.

New law and policy of remote sensing nations

India, the United Kingdom, and some other remote sensing nations have policies rather than laws. However, increasingly, major remote sensing nations are promulgating national laws. In 1999, Canada announced a policy that, in 2005, received Royal Assent and which came into force in April of 2007 as the *Remote Sensing Space Systems Act*. In 2007, the *German Act on Satellite Data Security* entered into force. These laws address the commercial availability of high-resolution imagery, and both seek to ensure national security interests within a commercial context. In 2008, Japan and France each passed a comprehensive national space law that includes sections on remote sensing. Court decisions in France, Germany, and the United States regarding the intellectual property and other aspects of remote sensing are also adding to the overall corpus of law.

Globalizing Earth observations

In the 1990s, the trend to internationalize Earth observation satellite operations began, and important new agreements were formulated. On November 19, 1998, the US National Oceanic and Atmospheric Administration and

the European Organisation for the Exploitation of Meteorological Satellites entered into an agreement on a joint polar-orbiting operational system, and a second agreement was entered into on June 24, 2003. *The International Charter on Space and Major Disasters* became operational on November 1, 2000. Some nations, like Belgium, do not have indigenous remote sensing capabilities but, nonetheless, are developing national remote sensing laws because they are participating in remote sensing consortia. It is clear that remote sensing law will continue to develop.

Bibliography

- Christol, C. Q., 1982. *The Modern International Law of Outer Space*. New York: Pergamon Press.
- Gabrynowicz, J., 2002a. *The United Nations Principles Relating to Remote Sensing of the Earth from Space: A legislative History – Interviews of Members of the United States Delegation*. Oxford, MS: The National Center for Remote Sensing, Air, and Space Law. ISBN 0-9720432-1-7.
- Gabrynowicz, J., 2002b. *Proceedings, The First International Conference on the State of Remote Sensing Law*. Oxford, MS: The National Center for Remote Sensing, Air, and Space Law. ISBN 0-9720432-3-3.
- Gabrynowicz, J., 2005. The perils of landsat from grassroots to globalization: a comprehensive review of U.S. remote sensing law with a few thoughts for the future. *Chicago Journal of International Law*, **6**, 72.
- Gabrynowicz, J., 2008. The second international conference on the state of remote sensing law. *Journal of Space Law*, **34**(1), ISSN 0095-7577 - 9720432-3-3, The National Center for Remote Sensing, Air, and Space Law: Oxford, MS.
- Graham, J. F., and Gabrynowicz, J., 2002. *Landsat 7: Past, Present and Future*. Oxford, MS: The National Center for Remote Sensing, Air, and Space Law. ISBN 0-9720432-0-9.
- The Land Remote Sensing Laws and Policies of National Governments: A Global Survey*, The National Center for Remote Sensing, Air, and Space Law, Oxford, MS, 2007. Available from <http://www.spacelaw.olemiss.edu/resources/pdfs/noaa.pdf>. Last accessed 21 June 2012.
- United Nations Office for Outer Space Affairs. *United Nations Office for Outer Space Affairs*, <http://www.unoosa.org/oosa/en/SpaceLaw/index.html>. Last accessed 21 June 2012.

LIDAR SYSTEMS

Robert Menzies
Jet Propulsion Laboratory, California Institute of
Technology, Pasadena, CA, USA

Synonyms

Ladar; Laser radar

Definitions

Lidar. Light Detection And Ranging
Ladar. LAser Detection And Ranging

Introduction

A lidar system in the strictly defined sense of the acronym measures range to a “target” that provides a signal that can be detected. Thus, the lidar system includes both a transmitter and a receiver. Ranging is accomplished using time-of-flight methods. The target can be a “hard” target that is essentially opaque to the lidar wavelength, not allowing measureable penetration beyond its range, or a “diffuse” scattering medium that allows penetration and range gating. Examples of the former are the surfaces of the Earth and other planets, or man-made objects. Examples of the latter are atmospheric aerosols and gases. In reality, these are terms that are commonly used but do not have strict, universally accepted definitions. In fact, the term “lidar” itself is commonly applied to systems that contain transmitters and receivers but do not have inherent range measurement capability. The lidar community is inclusive in this regard.

Following David Tratt’s introductory entry (“Emerging Technologies: Lidar”), which describes lidar basics and various classes or categories of lidar, we provide here a summary of the current capabilities in these various lidar applications areas. Our lidar categories are altimetry and mapping systems, backscatter systems, Doppler systems, and differential absorption systems. Comments on emerging technologies and methods are included. Lidar/ladar applications cover a wide range of activities and interests. The 3D imaging applications are a growth area with strong support from the defense community. System developments in this area are included only in brief overview mode. The balance in this entry is tilted more toward systems developed for scientific investigations.

Altimetry and mapping systems

Laser altimetry is relatively mature, with heritage in aircraft instruments, followed by Earth-orbiting, Mars-orbiting, and Lunar-orbiting systems. The early altimeter/mapping instruments used a form of threshold detection to trigger a circuit that enabled range measurement to a “first return” scattering surface. The implementation of fast waveform recovery, or multistop detection circuits, increases data rates but provides structure information in the line-of-sight dimension. The Geoscience Laser Altimeter System (GLAS) on the Earth-orbiting ICESat (Abshire et al., 2005) provided structure detail in the time domain, a capability that is essential for future use of laser 3D mappers in obtaining global estimates of biomass. High-resolution 3D imaging with very high depth resolution (~ 1 mm) can be achieved at km distances using fiber lasers and high bandwidth waveform encoding and decoding techniques (Buck et al., 2007). The current and next-generation systems combine multi-beam transmitter patterns with structural detail in the range dimension. The laser altimetric observational method provides line-of-sight detail that complements radar methods as well as higher spatial resolution in the cross dimensions. Spatial coverage is

a challenge, however. The advent of avalanche photodiode (APD) arrays and photon-counting receivers (e.g., Aull et al., 2002; Albota et al., 2002), combined with optical methods for simultaneous transmission of multiple beams, have greatly increased the mapping efficiencies of airborne and space systems. The use of statistical methods in a photon-counting mode has allowed the use of compact, high pulse-repetition frequency (prf), low pulse energy laser transmitters in various imaging and mapping systems (Degnan et al., 2008; Steinvall et al., 2008). The Lunar Orbiter Laser Altimeter (LOLA) instrument, scheduled to launched in June, 2009, uses a Diffractive Optical Element (DOE) to produce a 5-beam pattern for provision of more spatial coverage than with prior space laser altimeters (Ramos-Izquierdo et al., 2009; Smith et al., 2010). The DOEs have found use in various airborne laser 3D mappers.

An alternative to the use of scanners or elements such as DOEs, matched with APD arrays, is a flash lidar/ladar. The images in this type of system record the intensity reflected by the scene when flood-illuminated by the laser transmitter pulse. The laser transmitter irradiates the entire field of view of the receiver camera pixel array, and each pulse generates an entire frame of data (Stettner et al., 2005). The array elements are high-speed detectors that are periodically sampled in time at nanosecond timescales. The advances in hybridizing the focal planes with silicon CMOS read-out integrated circuits (ROICs), utilizing steady improvements in high-speed circuitry, provide the potential for growth with this approach. Laser sources can include semiconductor lasers and fiber lasers mated to power amplifiers.

Backscatter lidars

Here we include elastic backscatter lidars and various types of inelastic backscatter lidars (e.g., Raman, fluorescence). The emphasis is on atmospheric studies using these systems. The intensity or energy in the return signal is important with backscatter lidar measurements. Some method of calibration and/or normalization must be used in order to turn the data into useful observations. In the visible, the molecular density, if known sufficiently well, can be used to provide a Rayleigh backscatter intensity that effectively calibrates at least the range dependence of the lidar efficiency factor, or the efficiency factor itself at a particular atmospheric altitude where particle scattering is assumed negligible. This is not a viable technique at longer wavelengths in the infrared, due to the rapid decrease of the Rayleigh scattering cross section with increasing wavelength.

Backscatter lidars for cloud and aerosol studies date back to the early years of lidar, when ground-based lidars operating at visible wavelengths probed the stratospheric aerosol layers (e.g., Fiocco and Grams, 1964). The first Earth-orbiting lidar used for atmospheric studies was an elastic backscatter lidar (LITE, launched in 1994). GLAS operated both as an altimeter and an atmospheric lidar

(Spinhirne et al., 2005). Currently the CALIPSO lidar is in Earth orbit, being used for cloud and aerosol studies. The CALIPSO transmitter is a diode-array pumped Nd: YAG laser, by far the most commonly used in backscatter lidars. A wide dynamic range of pulse energies and pulse-repetition frequencies are available in this laser medium. The compact micropulse lidars, which emit pulses in the micro-Joule range, are deployed around the globe in networks such as the MPLNET (Micropulse Lidar Network) (Campbell et al., 2008). Cloud and aerosol detection, characterization, and monitoring algorithms continue to improve for these compact lidars, making them more useful for deployment. The underlying technologies are robust. The vertical profiling capabilities of these lidars cannot be duplicated with passive instruments. Recently, a compact backscatter lidar was deployed on the surface of Mars as part of the Phoenix mission (Whiteway et al., 2008).

A variant of the elastic backscatter lidar that is taking center stage in current and future atmospheric investigations is the High Spectral Resolution Lidar or HSRL. Although Doppler lidars are the ultimate high spectral resolution lidars, the term “HSRL” is commonly used in the lidar community to refer to a system that can separate the molecular Rayleigh backscatter signal from the aerosol backscatter signal. This obviates the need to assume a “lidar ratio” (i.e., aerosol extinction-to-backscatter ratio) when interpreting the range-dependent backscatter signals to deduce aerosol optical properties, thereby achieving more robust estimates of aerosol extinction coefficients. A progression in HSRL implementation has gone from early 1980s laser technology such as fragile dye laser systems (Shipley et al., 1983) to more robust solid-state lasers (Grund and Eloranta, 1991). Iodine vapor filters offer simplicity compared with the etalon filters in the HSRL receiver (Hair et al., 2001). More recently, airborne HSRL has been developed, and measurement results have been reported (Hair et al., 2008). The next-generation Earth-orbiting backscatter lidar for cloud and aerosol studies will likely be an HSRL. In fact, the European Space Agency’s Atmospheric Laser Doppler Instrument (ALADIN), a Doppler lidar in the Atmospheric Dynamics Mission (ADM) with atmospheric wind field measurements as its primary objective, is fundamentally an HSRL and will be used for investigations of aerosol optical properties (Ansmann et al., 2007) (see www.esa.int for further information).

Resonance fluorescence lidars have been in use for decades to study dynamics and thermal properties of the middle atmosphere, particularly the mesosphere. Lidars built to measure alkalis in the upper mesosphere were also used as Rayleigh backscatter lidars to measure density and temperature profiles in the stratosphere and mesosphere (Hauchecorne and Chanin, 1980). Developments in solid-state laser technology and injection seeding methods have resulted in systems that are more amenable to transportation and operation at remote sites

(e.g., She et al., 2007). Systems that interact with a variety of metals in addition to sodium and the other alkalis are now in development for investigations over a wider range of altitudes (Gardner, 2004).

Raman lidars are now commonly used for water vapor profiling and for characterizing the optical and microphysical properties of atmospheric aerosol. The latter method was described 20 years ago (Ansmann et al., 1990) and has continued to evolve into systems that are being used for characterization of major dust plumes that are transported long distances (e.g., Asian dust, Saharan dust) and for calibration/validation exercises (Mona et al., 2007). The former method has a long history and has slowly evolved with the use of improved techniques for minimizing background light, improved algorithms, and improved understanding of sources of bias. The use of Raman lidar for water vapor profiling in the lower atmosphere continues to gain credibility as the level of accuracy continues to improve (Adam and Venable, 2007; Leblanc and McDermid, 2008).

Doppler lidars

The atmospheric gas molecules and aerosol particles are in bulk motion in the dynamic atmosphere, and backscattering of laser radiation from the molecules and aerosol particles produces Doppler shifts in frequency. Doppler lidars detect these frequency shifts to deduce wind profiles. Two types of Doppler lidar have received attention over the years: direct detection and coherent detection lidars.

The coherent detection lidar is more sensitive and less difficult to implement at relatively longer wavelengths in the infrared, particularly at wavelengths longer than 1.5 μm , the so-called eye-safe region. The ultrahigh spectral resolution that is inherent with these systems makes coherent detection suitable for measuring Doppler-shifted backscatter from the atmospheric aerosol particles. The signal processing has similarities with Doppler radar. The use of rare-earth-doped solid-state laser technologies in the 2 μm wavelength region has been a popular choice for compact coherent detection systems. An example is the NOAA shipborne lidar, which has been used in many field campaigns (Tucker et al., 2009). Airborne systems date back to the mid-1980s when carbon dioxide gas laser transmitters were used (Bilbro et al., 1986). More recent, much more compact systems have also been deployed for measuring wind profiles with high spatial resolution (Hannon et al., 1999). Both the rare-earth-ion-doped solid-state crystal laser technology at 2 μm and the fiber laser technology developed primarily by the telecom industry have been employed in recent ground-based coherent Doppler lidars stationed at airports for airport safety enhancements. These lidar systems are being used for both wake vortex monitoring (e.g., Kopp et al., 2004) and wind shear detection and warning (e.g., Shun and Chan, 2008). Fiber laser

technologies are being incorporated into current and future systems.

Direct detection lidar is the appropriate choice for regions of the atmosphere containing very low aerosol particle concentration in the size range that is useful for optical scattering. The predominant scattering is molecular Rayleigh scattering. An early example was the use of direct detection Rayleigh lidar, modified with the incorporation of twin Fabry-Pérot interferometer filters in the receiver, for measurements of horizontal winds in the middle atmosphere (Chanin et al., 1989). An airborne direct detection Doppler lidar was developed, for tropospheric wind field measurements (Gentry et al., 2007). It is designed for autonomous operation on a high-altitude aircraft. The European Space Agency's ALADIN lidar is planned for launch in 2010, as the centerpiece instrument in the Atmospheric Dynamics Mission (ADM). ALADIN uses solid-state Nd: YAG laser transmitter technology, frequency-tripled to the 355 nm near-UV wavelength. It contains two receivers, one for the narrow-band Mie scattered radiation from the atmospheric aerosol particles (employing a multichannel Fizeau interferometer) and the other for the Rayleigh scattered radiation from the molecules (employing a double-edge Fabry-Pérot etalon). Accumulation CCD's are used in both receivers (see www.esa.int for further information).

Differential absorption lidars

Differential absorption lidars require typically two carefully selected closely spaced transmit wavelengths and a laser transmitter subsystem that has either discrete or continuous tunability in the desired spectral region to interact with the species of interest. Early systems used dye lasers or nonlinear optics such as optical parametric oscillators to provide tunability. Atmospheric ozone and water vapor have been favorite measurement subjects for decades. More recent systems rely on solid-state laser technologies and modern techniques for generating tunable single-mode radiation with high spectral purity. Airborne systems have progressed in sophistication, with corresponding reductions in mass and dimensions as well. The LASE (Lidar Atmospheric Sensing Experiment) system was demonstrated in the 1990s as an autonomous operation water vapor differential absorption lidar on the high-altitude ER-2 aircraft (Browell et al., 1997). Currently, intercomparison campaigns involving multiple airborne water vapor systems with different designs are being planned and implemented in order to better understand the accuracies of measurement and quantify biases that might exist (Behrendt et al., 2007). Results to date show that measurement accuracies are in good agreement with expectations.

Currently, a major challenge for differential absorption lidar is the measurement of atmospheric CO₂. Measurements with very high accuracy over regional to global scales would improve understanding of fluxes between

atmosphere, land surface, and ocean surface. The influence that increasing carbon dioxide mixing ratio has on climate change has spurred an interest in applying both passive and active remote sensing techniques to address this question. Desired mixing ratio accuracy levels of better than 1 % place great demands on a differential absorption lidar system itself and require the minimization of errors due to imperfect knowledge of the relevant atmospheric parameters (Menzies and Tratt, 2003; Ehret et al., 2008). Demonstrations of CO₂ mixing ratio measurement capability using ground-based, coherent detection systems have been reported (Gibert et al., 2008; Koch et al., 2008). Airborne systems are now being tested in flight campaigns, using both the solid-state 2 μ m laser technology and the 1.6 μ m fiber laser technology (Abshire et al., 2010; Spiers, et al., 2011). Studies of Earth-orbiting lidar systems for CO₂ measurements are being conducted under the sponsorship of European, US, and Japanese space agencies (ESA, NASA, and JAXA respectively).

Summary

Using an unofficial taxonomy of lidar systems, selected highlights of recent developments and future plans have been provided. Generally speaking, the future applications for altimetry and three-dimensional mapping will motivate increases in coverage within a given available time frame. This will most likely come from increases in total laser transmitter output power, along with optical technology. In other lidar application areas, engineering advances will be critical. For example, advances in compactness, electrical power efficiency, autonomy, and reliability will be essential for further use in hazard detection and monitoring, as well as expansion of regional and global networks for weather, climate, atmospheric composition, and environmental monitoring. Atmospheric greenhouse gas measurements, on a global scale, present high-precision measurement challenges. Nearly 50 years after the first demonstration of the laser, many lidar system applications are still driven by laser technology advances. For example, many applications still await the development of a wider range of laser sources in infrared spectral regions that are presently underutilized. The advent of the quantum cascade laser and other "bandgap-engineered" semiconductor laser technologies, as well as fiber laser/amplifier technologies, are good examples of continuing laser technology advances.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Abshire, J. B., Graham, H. R., Allan, R., Weaver, C. J., Mao, J., Sun, X., Hasselbrack, W. E., Kawa, S. R., and Biraud, S.,

2010. Pulsed airborne lidar measurements of atmospheric CO₂ column absorption. *Tellus*, **62B**, 770–783.
- Abshire, J. B., Sun, X., Riris, H., Sirota, J. M., McGarry, J. F., Palm, S., Yi, D., and Liiva, P., 2005. Geoscience laser altimeter system (GLAS) on the ICESat mission: on-orbit measurement performance. *Geophysical Research Letters*, **32**, L21S02, doi:10.1029/2005GL024028.
- Adam, M., and Venable, D., 2007. Systematic distortions in water vapor mixing ratio and aerosol scattering ratio from a Raman lidar. In *Proceedings of SPIE*, Vol. 6750, doi:10.1117/12.738205.
- Albota, M. A., Heinrichs, R. M., Kocher, D. G., Fouche, D. G., Player, B. E., O'Brien, M. E., Aull, B. F., Zayhowski, J. J., Mooney, J., Willard, B. C., and Carlson, R. R., 2002. Three-dimensional imaging laser radar with a photon-counting avalanche photodiode array and microchip laser. *Applied Optics*, **41**, 7671, doi:10.1364/AO.41.007671.
- Ansmann, A., Riebesell, M., and Weitkamp, C., 1990. Measurements of atmospheric aerosol extinction profiles with a Raman lidar. *Optics Letters*, **15**, 746–748.
- Ansmann, A., Wandinger, U., Le Rille, O., Lajas, D., and Straume, A. G., 2007. Particle backscatter and extinction profiling with the spaceborne high-spectral-resolution Doppler lidar ALADIN: methodology and simulations. *Applied Optics*, **46**, 6606–6622.
- Aull, B. F., Loomis, A. H., Young, D. J., Heinrichs, R. M., Felton, B. J., Daniels, P. J., and Landers, D. J., 2002. Geiger-mode avalanche photodiodes for three-dimensional imaging. *Lincoln Laboratory Journal*, **13**, 335–343.
- Behrendt, A., Wulfmeyer, V., Kiemle, C., Ehret, G., Flamant, C., Schaberl, T., Bauer, H.-S., Kiio, S., Ismail, S., Ferrare, R., Browell, E. V., and Whiteman, D. N., 2007. Intercomparison of water vapor data measured with lidar during IHOP-2002. Part II: airborne-to-airborne systems. *Journal of Atmospheric and Oceanic Technology*, **24**, 22–39, doi:10.1175/JTECH1925.1.
- Bilbro, J. W., DiMarzio, C. A., Fitzjarrald, D. E., Johnson, S. C., and Jones, W. D., 1986. Airborne Doppler lidar measurements. *Applied Optics*, **25**, 3952, doi:10.1364/AO.25.003952.
- Browell, E. V., Ismail, S., Hall, W. M., Moore, A. S., Jr., Kooi, S. A., Brackett, V. G., Clayton, M. B., Barrick, J. D. W., Schmidlin, F. J., Higdon, N. S., Melfi, S. H., and Whiteman, D. N., 1997. LASE validation experiment. In Ansmann, A., Neuber, R., Rairoux, P., and Wandinger, U. (eds.), *Advances in Atmospheric Remote Sensing with Lidar*. Berlin: Springer, pp. 289–295.
- Buck, J., Malm, A., Zakel, A., Krause, G., and Tiemann, B., 2007. High-resolution 3D coherent laser radar imaging. In *Proceedings of SPIE*, Vol. 6550, doi:10.1117/12.719483.
- Campbell, J. R., Sassen, K., and Welton, E. J., 2008. Elevated cloud and aerosol layer retrievals from micropulse lidar signal profiles. *Journal of Atmospheric and Oceanic Technology*, **25**, 685–700, doi:10.1175/2007JTECHA1034.1.
- Chanin, M. L., Garnier, A., Hauchecorne, A., and Porteneuve, J., 1989. A Doppler lidar for measuring winds in the middle atmosphere. *Geophysical Research Letters*, **16**, 1273–1276.
- Degnan, J., Machan, R., Leventhal, E., Lawrence, D., Jodor, G., and Field, C., 2008. In-flight performance of a second-generation, photon counting, 3D imaging lidar. In *Proceedings of SPIE*, Vol. 6950, Laser Radar Technology and Applications XIII, doi:10.1117/12.784759.
- Durand, Y., Chinal, E., Endemann, M., Meynart, R., Reitebuch, O., and Treichel, R., 2006. ALADIN airborne demonstrator: a Doppler wind lidar to prepare ESA's ADM-Aeolus Explorer mission. In *Proceedings of SPIE*, Vol. 6296, 62961D, doi:10.1117/12.680958.
- Ehret, G., Kiemle, C., Wirth, M., Amediek, A., Fix, A., and Houweling, S., 2008. Space-borne remote sensing of CO₂, CH₄, and N₂O by integrated path differential absorption lidar: a sensitivity analysis. *Applied Physics B*, **90**, 593–608, doi:10.1007/s00340-007-2892-3.
- Esselborn, M., Wirth, M., Fix, A., Tesche, M., and Ehret, G., 2008. Airborne high spectral resolution lidar for measuring aerosol extinction and backscatter coefficients. *Applied Optics*, **47**, 346–358.
- Fiocco, G., and Grams, G., 1964. Observation of the aerosol layer at 20 km by optical radar. *Journal of Atmospheric Science*, **21**, 323–324.
- Gardner, C. S., 2004. Performance capabilities of middle-atmosphere temperature lidars: comparison of Na, Fe, K, Ca, Ca⁺, and Rayleigh systems. *Applied Optics*, **43**, 4941–4956.
- Gentry, B., McGill, M., Schwemmer, G., Hardesty, M., Brewer, A., Wilkerson, T., Atlas, R., Sirota, M., Lindemann, S., and Hovis, F., 2007. Development of an airborne molecular direct detection Doppler lidar for tropospheric wind profiling. In *Proceedings of SPIE*, Vol. 6681, doi:10.1117/12.739379.
- Gibert, F., Flamant, P. H., Cuesta, J., and Bruneau, D., 2008. Vertical 2-μm heterodyne differential absorption lidar measurements of mean CO₂ mixing ratio in the troposphere. *Journal of Atmospheric and Oceanic Technology*, **25**, 1477–1497, doi:10.1175/2008JTECHA1070.1.
- Grund, C. J., and Eloranta, E. W., 1991. University of Wisconsin high spectral resolution lidar. *Optical Engineering*, **30**, 6–12.
- Hair, J. W., Caldwell, L. M., Krueger, D. A., and She, C.-Y., 2001. High-spectral-resolution lidar with iodine-vapor filters: measurements of atmospheric-state and aerosol profiles. *Applied Optics*, **40**, 5280–5294.
- Hair, J. W., Hostetler, C. A., Cook, A. L., Harper, D. B., Ferrare, R. A., Mack, T. L., Welch, W., Ramos-Izquierdo, L., and Hovis, F. E., 2008. Airborne high spectral resolution lidar for profiling aerosol optical properties. *Applied Optics*, **47**, 6734–6752.
- Hannon, S. M., Bagley, H. R., and Bogue, R. K., 1999. Airborne Doppler lidar turbulence detection: ACLAIM flight test results. In *Proceedings of SPIE*, Vol. 3707, 234, doi:10.1117/12.351378.
- Hauchecorne, A., and Chanin, M.-L., 1980. Density and temperature profiles obtained by lidar between 35 and 70 km. *Geophysical Research Letters*, **7**, 565–568.
- Koch, G. J., Petros, M., Barnes, B., Beyon, J. Y., Amzajerdian, F., Yu, J., Kavaya, M. J., and Singh, U. N., 2004. Validar: a testbed for advanced 2-micron Doppler lidar. In *Proceedings of SPIE*, Vol. 5412, pp. 87–98, doi:10.1117/12.542116.
- Koch, G. J., Beyon, J. Y., Gibert, F., Bernes, B. W., Ismail, S., Petros, M., Petzar, P. J., Yu, J., Modlin, E. A., Davis, K. J., and Singh, U. N., 2008. Side-line tunable laser transmitter for differential absorption lidar measurements of CO₂: design and application to atmospheric measurements. *Applied Optics*, **47**, 944–956.
- Kopp, F., Rahm, S., and Smalikho, I. M., 2004. Characterization of aircraft wake vortices by 2-μm pulsed Doppler lidar. *Journal of Atmospheric and Oceanic Technology*, **21**, 194–206.
- Leblanc, T., and McDermid, I. S., 2008. Accuracy of Raman lidar water vapor calibration and its applicability to long-term measurements. *Applied Optics*, **47**, 5592–5603.
- Menzies, R. T., and Tratt, D. M., 2003. Differential laser absorption spectrometry for global profiling of tropospheric carbon dioxide: selection of optimum sounding frequencies for high-precision measurements. *Applied Optics*, **42**, 6569, doi:10.1364/AO.42.006569.
- Mona, L., Amodeo, A., D'Amico, G., and Pappalardo, G., 2007. First comparisons between CNR-IAMM multi-wavelength Raman lidar measurements and CALIPSO measurements. In *Proceedings of the SPIE*, Vol. 6750, 675010, doi:10.1117/12.738011.

- Ramos-Izquierdo, L., Scott, V. S., Connelly, J., Schmidt, S., Mamakos, W., Guzek, J., Peters, C., Liiva, P., Rodriguez, M., Cavanaugh, H., and Riris, H., 2009. Optical system design and integration of the lunar orbiter laser altimeter. *Applied Optics*, **44**, 3035–3049, doi:10.1364/AO.44.007621.
- She, C. Y., Vance, J. D., Kawahara, T. D., Williams, B. P., and Wu, Q., 2007. A proposed all-solid-state transportable narrow-band sodium lidar for mesopause region temperature and horizontal wind measurement. *Canadian Journal of Physics*, **85**, 111–118.
- Shipley, S. T., Tracy, D. H., Eloranta, E. W., Trauger, J. T., Sroga, J. T., Roesler, F. L., and Weinman, J. A., 1983. High spectral resolution lidar to measure optical scattering properties of atmospheric aerosols. I: theory and instrumentation. *Applied Optics*, **22**, 3716–3724.
- Shun, C. M., and Chan, P. W., 2008. Applications of an infrared Doppler lidar in detection of wind shear. *Journal of Atmospheric and Oceanic Technology*, **25**, 637–655, doi:10.1175/2007JTECHA1057.1.
- Smith, D. E., Zuber, M. T., Neumann, G. A., Lemoine, F. G., Mazarico, E., Torrence, M. H., McGarry, J. F., Rowlands, D. D., Head, III, J. W., Duxbury, T. H., Aharonson, O., Lucey, P. G., Robinson, M. S., Barnouin, O. S., Cavanaugh, J. F., Sun, X., Liiva, P., Mao, D., Smith, J. C., and Bartels, A. E., 2010. Initial observations from the Lunar Orbiter Laser Altimeter (LOLA). *Geophysical Research Letters*, **37**, L18204, doi:10.1029/2010gl043751.
- Spiers, G. D., Menzies, R. T., Jacob, J., Christensen, L. E., Phillips, M. W., Choi, Y., and Browell, E. V., 2011. Atmospheric CO₂ measurements with a 2 μm airborne laser absorption spectrometer employing coherent detection. *Applied Optics*, **50**(14).
- Spinhirne, J. D., Palm, S. P., Hart, W. D., Hlavka, D. L., and Welton, E. J., 2005. Cloud and aerosol measurements from GLAS: overview and initial results. *Geophysical Research Letters*, **32**, L22S03, doi:10.1029/2005GL023507.
- Steinval, O., Sjoqvist, L., Henriksson, M., and Jonsson, P., 2008. High resolution lidar using time-correlated single-photon counting. In *Proceedings of SPIE*, Vol. 6950, doi:10.1117/12.778323.
- Stettner, R., Bailey, H., and Silverman, S., 2005. Large format time-of-flight focal plane detector development. In *Proceedings of SPIE*, Vol. 5791, pp. 288–292.
- Tucker, S. C., Brewer, W. A., Banta, R. M., Senff, C. J., Sandverg, S. P., Law, D. C., Weickmann, A. M., and Hardesty, R. M., 2009. Doppler lidar estimation of mixing height using turbulence, shear, and aerosol profiles. *Journal of Atmospheric and Oceanic Technology*, **26**, 673–688, doi:10.1175/2008JTECHA1157.1.
- Werner, C., Flamant, P. H., Reitebuch, O., Köpp, F., Streicher, J., Rahm, S., Nagel, E., Klier, M., and Herrmann, H., 2001. Wind infrared Doppler lidar instrument. *Optical Engineering*, **40**, 115, doi:10.1117/1.1335530.
- Whiteway, J., Daly, M., Carswell, A., Duck, T., Dickinson, C., Komguem, L., and Cook, C., 2008. Lidar on the Phoenix mission to Mars. *Journal of Geophysical Research*, **113**, E00A08, doi:10.1029/2007JE003002.

Cross-references

Cryosphere, Measurements and Applications
 Ocean, Measurements and Applications
 Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties
 Optical/Infrared, Radiative Transfer
 Optical/Infrared, Scattering by Aerosols and Hydrometeors

LIGHTNING

Rachel I. Albrecht¹, Daniel J. Cecil² and Steven J. Goodman³

¹Divisão de Satélites e Sistemas Ambientais (DSA/CPTEC), Instituto Nacional de Pesquisas Espaciais (INPE), Cachoeira Paulista, SP, Brazil

²Marshall Space Flight Center (MSFC), National Aeronautics and Space Administration (NASA), Huntsville, AL, USA

³National Environmental Satellite, Data, and Information Service (NESDIS), National Oceanic and Atmospheric Administration (NOAA), Silver Spring, MD, USA

Synonyms

Cloud flash; Ground flash

Definition

Lightning or *lightning discharge*. A series of transient and multiple electrical breakdown pulses producing high-current channels (Uman, 1987).

Lightning flash. A luminous manifestation accompanying a sudden electrical discharge which takes place from or inside a cloud or, less often, from high structures on the ground or from mountains (WMO, 2011).

Introduction

A lightning flash is a noncontinuous multi-scale physical process that ranges from the initial breakdown of air to the actual discharge propagation in discrete steps that can occur from cloud to ground (CG) or inside the clouds, i.e., intracloud (IC). In the case of CG lightning, the lightning channel formation is led by *stepped leaders* (that creates a conducting path between charge centers) and then followed by one or multiple *return strokes* that traverse the channel moving electric charges and neutralizing the leaders (Rakov and Uman, 2003). These series of return strokes are the lightning flash, and each stroke is guided by the *dart leaders* that propagate downward on the track of a preceding return stroke. CG flashes are also classified by the polarity of lowered charge: negative and positive. Negative flashes are more common and exhibit several return strokes, while positive flashes have a single or very few return strokes, but higher current than the negative ones. These processes occur too rapidly for the human eye to distinguish, and the flash appears as a single channel lasting for less than a second. Lightning detection networks typically look for the electric field changes associated with such processes. In the case of IC lightning, *recoil streamers* propagate within the track of positive branches of a bi-leader carrying strong negative charges. The lightning flash is terminated when the electric field is reduced to the point where it cannot sustain the discharge's propagation anymore.

The electromagnetic spectrum of lightning

The rapid release of electric energy inside the lightning channel generates a shock wave and electromagnetic radiation in a broad spectrum, see the entry [Radiation, Electromagnetic](#). The shock wave rapidly decays into an acoustic wave we know as thunder. The electromagnetic radiation ranges from radio frequencies through visible to X-rays and gamma rays, composing the basis for ground-based lightning location systems (LLS) and remote sensing from satellites.

Each lightning component (stepped leaders, return strokes, recoil streamers) emits electromagnetic energy proportional to the charge carried and its derivative in time. Negative stepped leaders are associated with strong negative currents in very short pulses (1 ms) and are detectable in very high frequencies (VHF, 1–200 MHz), as well as the dart leaders and recoil streamers, but with a relatively lower electrical current (MacGorman and Rust, 1998). The return strokes of a CG are high-energy discharges typically of the magnitude 10–100 kA in long pulses and radiate from the very low to high frequency range (1 kHz–10 MHz). Positive CG return strokes usually have continuous high current (>100 kA) and therefore are easily detected by VLF systems.

Radio emissions from lightning occur in the form of short pulses by accelerated charges during the fast-changing current steps, while the optical emissions occur from ionized and dissociated gases by thermal radiation of the lightning channel (Goodman et al., 1988). The heating in the channel reaches temperatures above 20,000 K resulting in optical emissions primarily in discrete atomic lines with some continuum at shorter wavelengths. Several measurements of lightning emission in the cloud top have shown strongest emissions at the neutral oxygen (OI(I)) and neutral nitrogen (NI(I)) lines, i.e., 777.4 and 868.3 nm in the near infrared, respectively (Goodman et al., 1988).

The radio electromagnetic waves of the lightning processes described above travel through the atmosphere and then are likely to be dissipated, reflected, scattered, refracted, and absorbed. The main effect is the dissipation, reducing the amplitude of the signal inversely proportional to the square of the distance. Ionospheric reflection, where the energy from waves with frequency lower than 5 MHz is trapped in the atmospheric waveguide formed by the ionosphere and the ground, permits long-range propagation of waves from high-energy return strokes. In the optical spectrum, the scattered energy by the cloud particles is observed from satellites as a diffuse light source at cloud top (Christian et al., 1989).

Ground-based lightning location systems

Several instruments can be used to locate lightning flashes, and more detail can be found at MacGorman and Rust, (1998) and Betz et al., (2009). The main technique consists of a network of sensors that detect IC and/or CG lightning by recording the electromagnetic radiation from

VLF to VHF continuously with time. The radiation detected by each sensor is then compared to other sensors in the network using two main location methods: the time-of-arrival and interferometer techniques. In the time-of-arrival method (TOA), time difference of lightning waveforms from several stations is computed and the location of lightning occurrence is given by the intersection of the hyperbolas for equal time differences. The interferometer method consists of determining the directions of the lightning waveform (azimuth and elevation) by analyzing the phase difference of an incident wave at several stations, and the intersection of these directions gives the location of the lightning source.

Today's operational lightning detection networks usually consist of different sensor types that use one or more location method for redundancy. These networks can be local, regional, or global depending on their operation baseline (distance between the sensors), and their detection efficiency and location accuracy are determined by the density of sensors and radio frequency used (Betz et al., 2009). Table 1 summarizes some of these more widely used lightning networks. The largest regional network is the US National Lightning Detection Network (NLDN) created in 1998, composed by 114 sensors operating in LF that locates mainly CG lightning in North America. Similar regional networks are found in Australia, Brazil, Canada, and Europe. Long-range networks operate in VLF and have been deployed worldwide in an attempt to locate lightning over remote areas like the oceans and the tropics. These networks operate with a sensor baseline of thousands of kilometers, which limits the detection efficiency to the stronger amplitude lightning signals (Cramer and Cummins, 1999).

Total lightning (IC + CG lightning) is monitored using VHF and a combination of LF and VLF or VHF and LF. In the USA, total lightning is monitored by several VHF Lightning Mapping Array (LMA) research networks (Table 1) developed by New Mexico Tech (Rison et al., 1999). The individual LMA regional networks consist of 10 or more stations extending ~80 km. The LMA measures the TOA of the magnetic peak signals at the different receiving stations to locate the source of impulsive VHF radio signals. Hundreds to thousands of sources per flash can be correlated in space and time, allowing a 3-D or 2-D lightning mapping of the channel over a regional domain of ~200 km.

Lightning detection from space

Several astronauts reported seeing lightning while looking down from space in the 1960s, describing flashes with hundreds of kilometers in extent and simultaneous flashes occurring between widely separated storms. Lightning was detected in early satellite imagery (Sparrow and Ney, 1971), and in 1981, the space shuttle astronauts recorded lightning in a 16 mm movie camera (Goodman et al., 1993). Although it was not their primary objective, several instruments onboard of the US Air Force DMSP

Lightning, Table 1 Ground-based lightning location systems operating in the world

Network	Frequency used	Type of discharges detected	Coverage area	Website
NLDN (US National Lightning Detection Network)	LF	Mainly CG	United States of America	http://www.vaisala.com/
CLDN (Canadian Lightning Detection Network)	LF	Mainly CG	Canada	http://www.ec.gc.ca/foudre-lightning/default.asp?lang=En&n=D88E34E8-1
EUCLID (European Cooperation for Lightning Detection)	LF	Mainly CG	Europe	http://www.euclid.org
RINDAT (Rede Integrada Nacional de Detecção de Descargas Atmosféricas)	LF	Mainly CG	South-Southeast Brazil	http://www.rindat.com.br/
LINET (LIghtning location NETwork)	VLF, LF	Total lightning (IC + CG)	Europe	http://www.pa.op.dlr.de/linet/
LDAR (Lightning Detection and Ranging)	VHF	Total lightning (IC + CG)	Florida, USA	http://branch.nsstc.nasa.gov/PUBLIC/LDARII/
LMA (Lightning Mapping Array)	VHF	Total lightning (IC + CG)	USA-New Mexico, Oklahoma, Northern Alabama, Western Texas, Colorado, Atlanta, Washington DC, Spain	http://lightning.nmt.edu/nmt_lms/
ENTLN (Earth Networks Total Lightning Networks)	ELF-HF	Total lightning (IC + CG)	Australia, Americas, Europe	http://www.earthnetworks.com/
STARNET (Sferics Timing and Ranging NETwork)	VLF	Mainly CG	Globe	http://www.zeus.iag.usp.br/
WWLLN (World Wide Lightning Location Network)	VLF	Mainly CG	South America and East Africa	http://wwlln.net/
Vaisala GLD360 (Global Lightning Dataset 360)	VLF	Mainly CG	Globe	http://www.vaisala.com/
GLN (Global Lightning Network)	VLF	Mainly CG	Globe	http://www.uspln.com/gln.html
ATDnet (Met Office's Arrival Time Difference network)	VLF	Mainly CG	Globe	http://www.metoffice.gov.uk/

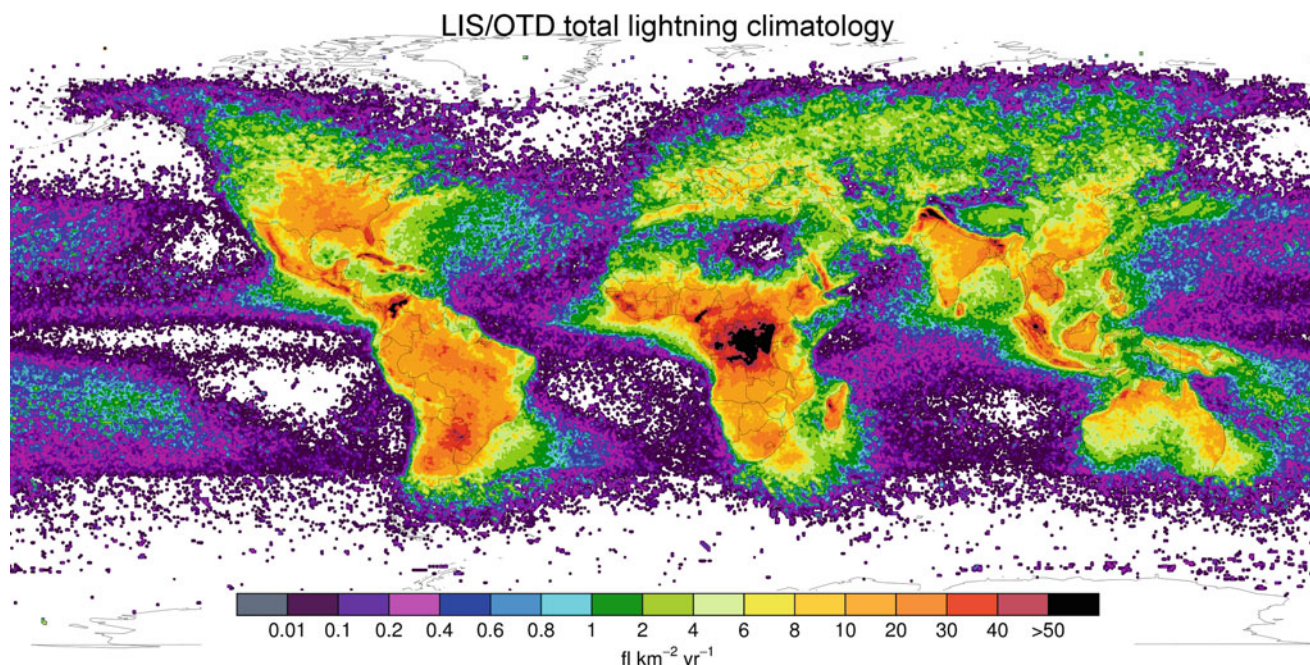
(Defense Meteorological Satellite Program) satellites have also recorded lightning, providing the first global lightning distribution map as a bonus to the mission (Goodman et al., 1993).

The Optical Transient Detector (OTD) onboard of the Microlab-1 (later renamed as OrbView-1) satellite was the first instrument designed to measure lightning from space day and night with storm scale resolution. The OTD operated between 1995 and 2000 in a 70° inclination low Earth orbit (see Low Earth Orbit (LEO)) at an altitude of 740 km. From this altitude, the OTD observed an individual storm for about 3 min. The design concept was based on the earlier research on optical emissions of lightning at cloud top (Christian and Goodman, 1987; Goodman et al., 1988). The OTD detected optical impulses with a 128×128 charge-coupled device (CCD) using a 1 nm narrow-band interference filter centered at 777.4 nm (Christian et al., 2003). Whereas the earlier satellite-based studies were limited to detecting visible lightning flashes during the darkness of night, the near-infrared wavelength combined with the use of spatial and temporal filtering used by OTD also allowed lightning detection during daylight. In 1997, the Lightning Imaging Sensor (LIS) onboard the Tropical Rainfall

Measuring Mission (TRMM) (Kummerow et al., 1998) was launched into a lower orbit inclination of 35° at an altitude of 350 km, later raised to 402 km in August 2001 to extend the mission lifetime. From this altitude, the LIS observed an individual storm for about 90 s.

The OTD was a flight qualified engineering model of the LIS, and thus, they share the same basic design heritage. In both OTD and LIS, the signal is read out from the focal plane into a real-time event processor for lightning event detection. The background scene is updated during each frame readout sequence and when a pixel's brightness compared to the prior background values exceeds a threshold, it is identified as a lightning *event*. The events are sent to the satellite ground station for geolocation processing in space and time, and an algorithm clusters the events into "flashes" (multiple CCD events grouped into time and space). The flash cannot be distinguished between CG and IC lightning, although in a statistical sense, the fraction of CG and IC flashes might be retrievable from a large sample of flashes (Koshak, 2010).

FORTE (Fast On-Orbit Recording of Transient Events) satellite was built by Los Alamos National Laboratory to study lightning signals from space (Jacobson et al., 2000,



Lightning, Figure 1 Total lightning climatology derived from OTD (1995–2000) and LIS (1998–2010) at 0.5° resolution.

Suszcynsky et al., 2000; Hamlin et al., 2009). Launched in August 1997 at 800 km of altitude with a 70° of inclination and a circular orbit, the optical lightning location system has the same design of OTD and LIS but it also carries a broad band photo diode and the VHF receivers had two broad band channels which are selectable from a grid covering the entire high frequency and to very high frequency (Hamlin et al., 2009), allowing a combined optical and radio frequency observations of lightning. FORTE demonstrated that lightning can be located from space based on multiple-satellite VHF receivers. The “ORAGE” project has also been studying the possibility of locating lightning flashes using VHF-UHF interferometry from a constellation of microsatellites (Bondiou-Clergerie et al., 1999).

OTD and LIS findings

The first global distribution of total lightning was derived from 5 years of OTD measurements by Christian et al., (2003), who found that the annual average global flash rate is 44 fl s⁻¹, with a maximum of 55 fl s⁻¹ in the boreal summer and a minimum of 35 fl s⁻¹ in the boreal winter. Recently, Blakeslee et al. (2012) and Cecil et al. (2012) found that these values remained nearly the same combining OTD (1995–2000) and LIS observations (1998–2010). These authors also showed that all continents display a strong diurnal variation with lightning peaking in the late afternoon, while oceans exhibit a

minimal nearly flat diurnal variation, but morning hours are typically slightly enhanced over afternoon. In Figure 1, we present the updated LIS/OTD climatology for 16 years of OTD (1995–2000) and LIS (1999–2010) combined observations of total lightning flash rate density (FRD, fl km⁻² year⁻¹) from Marshall Space Flight Center gridded LIS-OTD climatology product (High Resolution Flash Climatology, HRFC_COM_FR - Cecil et al. 2012). The difference between land and ocean can be clearly observed, with lightning occurring more frequently over continental (> 20 fl km⁻² year⁻¹) regions having greater instability and stronger vertical motion than oceanic environments. However, some coastal regions presented moderate FRD (1–10 fl km⁻² year⁻¹) associated with frequent synoptic scale extratropical cyclones and cold fronts (such as south-southeast coasts of Brazil, South Africa, Australia, and United States), and large-scale convergence zones (such as the South Atlantic, South Pacific, and the Intertropical Convergence Zones).

High elevated and complex terrain regions over the tropics can be identified by high thunderstorm activity (> 30 fl km⁻² year⁻¹) at the mountains foot (e.g., Andes, Himalayas, Sierra Madre Occidental, Cameroon Line, and Mitumba Mountains). Congo Basin is dramatically highlighted by its extensive area of large FRD (> 50 fl km⁻² year⁻¹), where the greatest annual number of individual thunderstorms is observed (Zipser et al., 2006). However, higher resolution (0.10°) LIS climatological maps highlighting topographical features and complex

terrain indicate the Congo Basin has the second highest climatological FRD, with $232 \text{ fl km}^{-2} \text{ year}^{-1}$ (at the foothills of Mitumba Mountains), while the total lightning “hot spot” on Earth is observed over Lake Maracaibo with $250 \text{ fl km}^{-2} \text{ year}^{-1}$ (Albrecht et al., 2011). Lake Maracaibo thunderstorm activity is very localized, determined by nocturnal convergence of land-lake and mountain-valley breezes over a warm lake, building the perfect scenario for thunderstorm development of more than 300 days per year (Albrecht et al., 2011). Frequent lightning activity ($30\text{--}50 \text{ fl km}^{-2} \text{ year}^{-1}$) is also observed over Florida, Cuba, and Indonesia-Malaysia due to land-ocean sea breezes and over the borders of Argentina, Paraguay, and Brazil where the greatest individual flash rates per mesoscale convective systems are observed (Cecil et al., 2005; Zipser et al., 2006).

In addition to mapping the lightning distribution, the instrument suite on the TRMM satellite allows more detailed characterization of the thunderstorms producing lightning. The TRMM radar and radiometer (see [Micro-wave Radiometers](#)) also show more intense storms over land (Cecil et al., 2005; Zipser et al., 2006). But for a given radar or radiometer signature, a storm over land is likely to produce more lightning than an otherwise similar storm over ocean (Liu et al., 2011). This suggests differences in the mixed phase microphysics and precipitation, hinted at by lightning but not resolved by the radar or radiometer uniquely by themselves.

More information on LIS and OTD can be found at <http://thunder.msfc.nasa.gov/>.

The future of lightning mapping from space

The next generation of NOAA Geostationary Operational Environmental Satellite (GOES-R) series and the EUMETSAT Meteosat Third Generation (MTG) will detect, locate, and measure continuous total lightning activity over their full disk with a nominal resolution of 10 km. GOES-R will carry the Geostationary Lightning Mapper (GLM) and it is scheduled to be launched in late 2015, while the MTG will carry the Lightning Imager (LI) and it is scheduled to be launched in 2018. Both GLM and LI are heritages of OTD and LIS, but GOES-R and MTG are equipped with improved communications systems and much greater telemetry bandwidth to ensure a continuous and reliable flow of the remote sensing products. The GOES-R series will maintain the 2-satellite system over the western hemisphere, with the operational GOES-R satellites at 75°W and 137°W . The GLM and LI together will provide continuous full-disk total lightning for storm warning and nowcasting (e.g., early warnings of tornadic activity, hail, and floods – see [Severe Storms](#)) for half of the globe. A geostationary lightning imager (GLI) having more limited coverage of mainland China and adjacent ocean is also planned for the Chinese FY-4 next-generation geostationary satellite series. More information on GOES-R GLM and MTG-LI can be found at <http://www.goes-r.gov/> and <http://www.eumetsat.int/>, respectively.

Bibliography

- Albrecht, R. I., et al., 2011. The 13 years of TRMM lightning imaging sensor: from individual flash characteristics to decadal tendencies. In *Proceedings XIV International Conference on Atmospheric Electricity*. Rio de Janeiro, 08–12, Aug 2011.
- Betz, H. D., Schumann, U., and Laroche, P., 2009. *Lightning: Principles, Instruments and Applications. Review of Modern Lightning Research*. Dordrecht: Springer, p. 641.
- Blakeslee, R. J., Mach, D. M., Bateman, M. G., and Bailey, J. C., 2012. Seasonal variations in the lightning diurnal cycle and implications for the global electric circuit. *Atmospheric Research*, <http://dx.doi.org/10.1016/j.atmosres.2012.09.023>.
- Bondiou-Clergerie, A., Blanchet, P., Th  ry, C., Delannoy, A., Lojou, J.Y., Soulage, A., Richard, P., Roux, F., and Chauzy, S., 1999. “ORAGES”: a project for space-borne detection of lightning flashes using interferometry in the VHF-UHF band. In *Proceedings of the 11th International Conference on Atmospheric Electricity*, Guntersville, Alabama, pp. 184–187.
- Cecil, D. J., Buechler, D. E., and Blakeslee, R. J., 2012. Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. *Atmospheric Research*, doi: <http://dx.doi.org/10.1016/j.atmosres.2012.06.028>.
- Cecil, D. J., Goodman, S. J., Boccippio, D. J., Zipser, E. J., and Nesbitt, S. W., 2005. Three years of TRMM precipitation features. Part I: radar, radiometric, and lightning characteristics. *Monthly Weather Review*, **133**(3), 543–566.
- Christian, H. J., and Goodman, S. J., 1987. Optical observations of lightning from a high altitude airplane. *Journal of Atmospheric and Oceanic Technology*, **4**, 701–711.
- Christian, H. J., Blakeslee, R. J., and Goodman, S. J., 1989. The detection of lightning from geostationary orbit. *Journal of Geophysical Research*, **94**, 13329–13337.
- Christian, H. J., et al., 2003. Global frequency and distribution of lightning as observed from space by the optical transient detector. *Journal of Geophysical Research*, **108**, 4005, doi:10.1029/2002JD002347.
- Cramer, J. A., and Cummins, K. L., 1999. Long-range and trans-oceanic lightning detection. In *Proceedings 11th International Conference on Atmospheric Electricity*. Guntersville, 7–11, June 1999, pp. 250–253.
- Goodman, S. J., Christian, H. J., and Rust, W. D., 1988. Optical pulse characteristics of intracloud and cloud-to-ground lightning observed from above clouds. *Journal of Applied Meteorology*, **27**, 1369–1381.
- Goodman, S. J., Christian, H. J., and Rust, W. D., 1993. Global observations of lightning. In Gurney, R. J., Foster, J. L., and Parkinson, C. L. (eds.), *Atlas of Satellite Observations Related to Global Change*. Cambridge, UK: Cambridge University Press, pp. 191–219.
- Hamlin, T., Wiens, K. C., Jacobson, A. R., Light, T. E. L., and Eack, K. B., 2009. Space- and ground-based studies of lightning signatures. In *Lightning: Principles, Instruments and Applications. Review of Modern Lightning Research*. Dordrecht: Springer, pp. 287–2017.
- Jacobson, A. R., Cummins, K. L., Carter, M., Klingner, P., Dupre, D. R., and Knox, S. O., 2000. FORTE radiofrequency observations of lightning strokes detected by the National Lightning Detection Network. *Journal of Geophysical Research*, **105**, 15653–15662.
- Koshak, W. J., 2010. Optical characteristics of OTD flashes and the implementations for flash-type discrimination. *Journal of Atmospheric and Oceanic Technology*, **27**, 1822–1838.
- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J., 1998. The tropical rainfall measuring mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, **15**, 809–817.

- Liu, C., Cecil, D., and Zipser, E. J., 2011. Relationships between lightning flash rates and passive microwave brightness temperatures at 85 and 37 GHz over the tropics and subtropics. *Journal of Geophysical Research*, **116**, D23108.
- MacGorman, D. R., and Rust, W. D., 1998. *The Electrical Nature of Storms*. New York: Oxford University Press.
- Rakov, V. A., and Uman, M. A., 2003. *Lightning: Physics and Effects*. Cambridge, UK: Cambridge University Press, p. 687.
- Rison, W., Thomas, R. J., Krehbiel, P. R., Hamlin, T., and Harlin, J., 1999. A GPS-based three-dimensional lightning mapping system: initial observations in central New Mexico. *Geophysical Research Letters*, **26**, 3573–3576.
- Sparrow, J. G., and Ney, E. P., 1971. Lightning observations by satellite. *Nature*, **232**, 540–541.
- Suszcynsky, D. M., Kirkland, M. W., Jacobson, A. R., Franz, R. C., Knox, S. O., Guillen, J. L. L., and Green, J. L., 2000. FORTE observations of simultaneous VHF and optical emissions from lightning: Basic phenomenology. *Journal of Geophysical Research*, **105**(D2), 2191–2201.
- Uman, M. A., 1987. *The Lightning Discharge*. New York: Elsevier, p. 228.
- World Meteorological Organization (WMO), 2011. METOTERM http://www.wmo.int/pages/prog/lsp/meteoterm_wmo_en.html. Visited on 16, Oct 2011.
- Zipser, E., Cecil, D., Liu, C., Nesbitt, S. W., and Yorty, S., 2006. Where are the most intense thunderstorms on Earth? *Bulletin of the American Meteorological Society*, **87**, 1057–1071.

Cross-references

Microwave Radiometers
Optical/Infrared, Radiative Transfer
Radiation, Electromagnetic
Severe Storms

LIMB SOUNDING, ATMOSPHERIC

Nathaniel Livesey
Jet Propulsion Laboratory, California Institute of
Technology, Pasadena, CA, USA

Synonyms

Limb profiling; Occultation measurements

Definition

Limb. The portion of a planetary (or stellar) atmosphere at the outer boundary of the disk, viewed “edge on.”

Limb sounding. Atmospheric remote sounding technique involving observing radiation emitted or scattered from the limb.

Occultation. Atmospheric remote sounding technique involving observing radiation emitted (or reflected) by a distant body (solar, stellar, lunar, or an orbiting satellite), transmitted along a limb path through an absorbing and/or scattering planetary atmosphere, and detected by a remote observer.

Introduction

Limb sounding is a widely used atmospheric remote sounding technique, whereby the atmosphere is viewed “edge on” by a space- or airborne instrument. Limb sounding observations are made from the microwave and infrared – where thermal emission is observed – to the visible and ultraviolet, where observations are typically of sunlight scattered in the limb or of airglow. A wide range of spaceborne limb sounding instruments have been used to observe atmospheric temperature, composition, and dynamics from the upper troposphere (~10 km altitude) to the mid-thermosphere (~450 km). A closely related technique, known as “occultation,” involves observing the atmospheric absorption and/or scattering of radiation emitted by a remote source (solar, lunar, stellar, or, more recently a GPS satellite).

Limb sounding has significant advantages over nadir sounding (i.e., viewing straight down) or near-nadir sounding. Firstly, scanning the instrument field of view vertically across the atmospheric limb can give atmospheric profile information with greater vertical resolution than is typically possible from nadir sounders. In addition, complexities associated with emission or reflection of radiation by the planetary surface can be avoided. Finally, by viewing a significantly longer atmospheric path than nadir sounders, limb viewing instruments can achieve a stronger signal to noise for observations of tenuous atmospheric trace gases. However, this same long path length (typically a few 100 km) results in a poorer horizontal resolution than is possible with nadir sounding instruments.

With the exception of the infrared Mars Climate Sounder instrument (MCS, McCleese et al., 2007) on the Mars Climate Orbiter, limb sounding observations have been confined to those of Earth’s atmosphere and are the focus of the discussion in this entry.

Principles and techniques

Limb radiances and line broadening

Each limb view is associated with a particular “tangent height” – the closest distance from the limb ray to the Earth’s surface. High tangent height views typically give small signals, due to the tenuous atmosphere at these altitudes. As tangent altitudes decrease, atmospheric emission or scattering strengthens, increasing the observed signals. Eventually, the atmosphere becomes sufficiently opaque that signals from lower regions in the atmosphere are absorbed by the layers above and not seen by the instrument. At this point, radiances tend to remain fairly constant with decreasing tangent altitude (or to change only slightly, due to second-order geometrical effects) and are said to be “saturated” or “blacked out,” as the signal continues to derive largely from the lowermost nonopaque layers. Refraction is significant for limb rays in the lower atmosphere but is generally negligible above ~20 km.

Molecular spectral lines are broadened in the atmosphere by a combination of the ensemble of Doppler shifts from the thermal motion of the molecules (“Doppler” broadening) and by collisions with other molecules (“collision” or “pressure” broadening). The latter generally dominates line widths of infrared and microwave signals up to ~ 60 km, while for visible and ultraviolet wavelengths, Doppler broadening dominates throughout the bulk of the atmosphere. Pressure broadening of spectral lines can provide valuable information on the vertical distribution of trace gases (in addition to the information gained by vertically scanning the instrument field of view) with frequencies further from line centers conveying information on lower regions of the atmosphere, where lines are broad enough to contribute to the observed signals. For wavelengths where pressure broadening is insignificant, vertical distribution information can still be obtained by observing in multiple spectral regions having different atmospheric absorptions (and thus penetration depths).

Solar occultation and related observations

Observation of the atmospheric absorption of solar radiation (“direct sun” measurements) has a long heritage in atmospheric science (e.g., the observations of ozone pioneered in the 1920s by Dobson). Solar occultation is a natural extension of these ground-based techniques (and similar observations from balloon and aircraft vantage points). An instrument on a low Earth-orbiting spacecraft can perform an occultation observation during sunrise and sunset on each of ~ 14 orbits per 24 h period. Typically, occultation instruments observe a narrow portion of the solar disk and track this as it rises or sets through the atmosphere. The strong solar signal provides excellent signal to noise and obviates the need to cool the instrument or its detectors. The observations of the sun above the atmosphere, before sunset or after sunrise, can be used to ensure a stable instrument calibration.

A succession of solar occultation instruments have provided a long record of atmospheric composition observations including the Stratospheric Aerosol and Gas Experiment (SAGE) I, II, and III series of instruments (McCormick et al., 1989), the Polar Ozone and Aerosol Measurement (POAM) instruments (Lucke et al., 1999), and the Halogen Occultation Experiment (HALOE) on the Upper Atmosphere Research Satellite (UARS) (Russell et al., 1993). Occultation observations were also made by the Atmospheric Trace Molecule Spectroscopy (ATMOS) instrument flown on the Space Shuttle ATLAS program (Gunson et al., 1996), and the Improved Limb Atmospheric Spectrometer instruments (ILAS I and II) on the Japanese Advanced Earth Observing Satellites (ADEOS I and II). The Scanning Imaging Absorption Spectrometer for Atmospheric CHartography instrument (SCIAMACHY, Bovensman et al., 1999) on the European Envisat performs solar occultation measurements in addition to limb and nadir imaging. Most recently, the

Atmospheric Chemistry Experiment’s Fourier Transform Spectrometer and Measurement of Aerosol Extinction in the Stratosphere and Troposphere Retrieved by Occultation (ACE/FTS and ACE/MAESTRO) instruments have been continuing and augmenting this record (Bernath et al., 2005). Currently, ACE is the only operating mission employing solar occultation, though other concepts are in formulation including plans to fly a SAGE III instrument on the International Space Station.

While offering good vertical resolution and outstanding signal to noise and calibration stability, solar occultation instruments are fundamentally limited by orbital geometry to making only ~ 30 observations per 24 h period. While some instruments augment this coverage with observations of lunar or stellar occultations, these, by definition, have poorer signal to noise than the solar occultation observations.

As described below, limb sounding observations of atmospheric emission or of scattered solar radiation offer comparable vertical resolution to occultation but have the advantage that observations can be made on a near-global basis daily.

Radio occultation

Observations of the atmospheric occultation of signals broadcast by GPS are a more recent development. In this technique, observations of refractive phase shift, as opposed to atmospheric absorption in different spectral regions, form the basis for the measurement. GPS occultation yields information on atmospheric refraction, and in turn temperature and/or water vapor profiles. More information on this technique is given elsewhere in this volume.

Microwave limb sounding

Atmospheric microwave emissions are generally associated with molecular rotational transitions, theoretically enabling observation of any atmospheric species with a significant dipole moment. Microwave limb sounding instruments have made observations of a wealth of species in the frequency range from ~ 60 GHz (5 mm wavelength) to ~ 2.5 THz (120 μm). Microwave signals are unaffected by aerosols and all but the thickest clouds, as the wavelengths used are longer than the typical particle sizes. This enables microwave observations of atmospheric composition in a limb sounding geometry in regions that are too cloudy for observations at other wavelengths.

To date, five spaceborne instruments employing limb sounding at microwave frequencies have flown: The Microwave Atmospheric Sounder (MAS) as part of the ATLAS payload on the Space Shuttle (Croskey et al., 1992), the Microwave Limb Sounder (MLS) instruments on the NASA UARS and Aura satellites (Barath et al., 1993; Waters et al., 2006), the Submillimeter Radiometer (SMR) on the Swedish Odin satellite (Murtagh et al., 2002), and, most recently, the Submillimeter-Wave Limb

Emission Sounder (SMILES) on the International Space Station (Ozeki et al., 2001).

Microwave instruments can achieve arbitrarily fine frequency resolutions, enabling individual transition lines to be resolved in great detail. The observations of line shape enable simultaneous observation of both atmospheric pressure (largely affecting the width of a given transition line in the pressure-broadening regime) and species abundance (largely affecting the line strength). By combining inferred atmospheric pressure information with limb tangent altitude information, and assuming hydrostatic balance, atmospheric temperature profile information can be obtained.

The field-of-view width for a microwave instrument is determined by the antenna size and wavelength employed, with narrower fields of view achieved for larger antennae and/or shorter wavelengths and somewhat large antennae dictated for many observations. For example, the Aura MLS instrument's 1.6 m antenna has a field of view that is ~ 3.5 km wide at the limb (full width, half maximum) at 200 GHz from a 700 km orbit. The lower vertical range of microwave limb sounding instruments is limited by continuum absorption from oxygen, nitrogen, and water vapor, with ~ 8 km altitude typically being the deepest penetration.

Infrared limb sounding

As with microwave limb sounding, infrared instruments observe thermal emission from the atmosphere, in this case mostly arising from molecular vibrational transitions. Again, collisional broadening enables determination of atmospheric pressure at the tangent point. Although not all infrared limb sounding instruments have the spectral resolution to resolve individual line shapes, pressure information can generally still be obtained from broader-band measurements.

Scattering and emission from clouds pose a more significant limitation to infrared limb sounders than microwave instruments, particularly in the tropics where clouds are prevalent in the upper troposphere. In clear-sky regions, infrared limb sounders can typically penetrate a few kilometers deeper than microwave sounders, but continuum absorption is, again, the ultimate limitation to this penetration. Infrared instruments can more easily achieve narrower fields of view than those in the microwave, and this can translate into a finer vertical resolution for the geophysical observations. However, the detectors typically need to be cooled (e.g., to ~ 70 K) in order to achieve a scientifically useful signal to noise.

Infrared limb sounding instruments have a long history in atmospheric science, starting with the Limb Radiance Inversion Radiometer (LRIR) on Nimbus 6, followed by the Limb Infrared Monitor of the Stratosphere (LIMS, Gille and Russell, 1984) and Stratospheric and Mesospheric Sounder (SAMS, Drummond et al., 1980) instruments on Nimbus 7. UARS included two infrared limb sounding instruments – the Cryogenic Limb Array

Etalon Spectrometer (CLAES, Roche et al., 1993) and the Improved Stratospheric and Mesospheric Sounder (ISAMS, Taylor et al., 1993). More recent infrared limb sounders include the Michelson Interferometer for Passive Atmospheric Sounding instrument (MIPAS, Fischer et al., 2008) on ESA's Envisat spacecraft and the High-Resolution Dynamics Limb Sounder (HIRDLS, Gille et al., 2008) on NASA's Aura satellite.

The Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) instrument on the Thermosphere, Ionosphere, Mesosphere Energetics and Dynamics (TIMED) mission (Russell et al., 1999) is another recently launched infrared limb sounder mainly focusing on the chemistry and structure of the upper atmosphere.

Visible and ultraviolet limb sounding

Limb viewing instruments at visible and ultraviolet wavelengths generally observe sunlight scattered by the atmospheric limb. These include the Optical Spectrograph and Infrared Imaging System (OSIRIS, Llewellyn et al., 2003) instrument on Odin and the planned limb sensor for the Ozone Mapping and Profiling Suite (OMPS) on the NPOESS (National Polar-orbiting Operational Environmental Satellite System) Preparatory Project (NPP). The SCIAMACY instrument on Envisat also includes a limb scattering capability. As collisional broadening does not significantly affect the (mainly electronic or vibronic) molecular transitions at these wavelengths, tangent pressure cannot be deduced from the observations, and the height registration of the resulting geophysical products is more critically reliant on independent knowledge of spacecraft pointing than is the case for longer wavelength observations.

In addition to limb scattering sounders, past instruments have observed visible atmospheric airglow emissions in the upper atmosphere. These include the Wind Imaging Interferometer (WINDII, Shepherd et al., 1993) and the High-Resolution Doppler Imager (HRDI, Hays et al., 1993) instruments, both on UARS, which used these observations to deduce upper atmospheric dynamics.

Inversion approaches for limb sounding instruments

Although scanning the field of view of an instrument up and down the atmospheric limb enables high-resolution observations of vertical profiles, the observed signals are (as with nadir sounding) affected by emission, absorption, and scattering throughout the ray path. Disentangling the impact of each atmospheric layer on the observed signal and deducing vertical profiles of temperature and composition is a nontrivial task, commonly known as a "retrieval" or "inverse" calculation.

A variety of techniques have been employed for limb sounding retrievals. The so-called onion peeling approach uses observations at the top of the limb scan to characterize the uppermost atmospheric region. This is then

accounted for when characterizing the next layer down, using lower-altitude limb views, and so on. A drawback of this technique is that the resulting profile depends strongly on the knowledge of the uppermost regions, where signal to noise is typically poor.

The most commonly adopted approach for limb sounding retrievals is the well-established “optimal estimation” method (Rodgers, 2000), which seeks the atmospheric state that matches all the observed signals simultaneously (taking into account potential noise on each signal). Although more computationally intensive than simpler approaches, this need not be a barrier with modern computing capabilities. Indeed, the most computationally demanding part of the calculation is typically the “forward model” (the computation that estimates the signal that would be observed by the instrument for a given atmospheric state), which is a central part of all but the simplest retrieval approaches, and upon which the accuracy of the resulting geophysical profiles ultimately depends.

Inhomogeneity along the limb line of sight can introduce biases in limb sounding retrievals, particularly in regions of strong atmospheric gradients. Some retrieval methods employ an iterative approach, whereby horizontal gradient information from a first pass is considered in a later retrieval step. In cases where the instrument line of sight is aligned with the spacecraft velocity vector, successive limb scans take multiple views through the same region of atmosphere, enabling a “tomographic” approach to the retrieval calculation to be taken (e.g., Livesey and Read (2000)).

Notable findings from limb sounding observations

The near-global daily coverage and good vertical resolution of limb sounding instruments has provided a wealth of information on atmospheric structure and composition from the upper troposphere through to the thermosphere. The early observations from LIMS and SAMS set the stage, with zonal-mean information on the abundance of key stratospheric and mesospheric trace gases. The three atmospheric composition limb sounders (CLAES, ISAMS, and MLS) on UARS, along with the HALOE solar occultation instrument, provided valuable insights into the dynamics and chemistry of Earth’s stratosphere, most notably processes associated with chemical **stratospheric ozone** loss and the transport of air into and throughout the stratosphere. UARS observations also provided valuable information on the impact of volcanic gases and the resulting aerosols on the stratosphere, following the dramatic June 1991 eruption of Mt. Pinatubo in the Philippines (4 months before the UARS launch).

In addition to stratospheric and mesospheric composition observations, UARS MLS provided unprecedented information on water vapor and ice clouds in the upper troposphere. The Aura MLS and HIRDLS instruments have enhanced this record providing the first daily global observations of upper tropospheric ozone, carbon

monoxide (MLS only), and nitric acid. The upper troposphere is an important region of the atmosphere for climate, as it is where water vapor (the strongest greenhouse gas) and ozone have their largest radiative impact.

Outlook

At the time of writing, the only limb sounding instruments in operation are Aura MLS, Odin SMR and OSIRS, and Envisat MIPAS. The SMILES instrument experienced a critical failure after ~6 months of operation, although plans are in formulation for a possible fix. While several new limb sounding instrument concepts are under formulation, to date none have been confirmed for launch. The most mature is the ESA Process Exploration through Measurements of Infrared and millimeter-wave Emitted Radiation (PREMIER) mission, which includes infrared and microwave limb sounding instruments observing the upper troposphere and lower stratosphere.

Conclusion

Limb sounding instruments provide a wealth of information on the composition, structure, and dynamics of Earth’s atmosphere, through observations of emitted or scattered radiation in an “edge on” viewing geometry. Limb sounding offers a valuable combination of good vertical resolution and near-global daily coverage, using wavelengths ranging from the microwave to the ultraviolet, and can provide observations from the upper troposphere to the middle thermosphere. Limb sounding observations have led to important discoveries concerning key dynamical and chemical processes in Earth’s stratosphere (including those processes associated with the “ozone hole”), and in the upper troposphere where water vapor and ozone have their strongest greenhouse forcing.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Barath, F. T., et al., 1993. The upper atmosphere research satellite microwave limb sounder experiment. *Journal of Geophysical Research*, **98**, 10, 751–10, 762.
- Bernath, P. D., et al., 2005. Atmospheric chemistry experiment (ACE): mission overview. *Geophysical Research Letters*, **32**, L15S01, doi:10.1029/2005GL022386.
- Bovensman, H., Burrows, J. P., Buckwitz, B., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H., 1999. SCIAMACHY: mission objectives and measurement modes. *Journal of Atmospheric Science*, **56**, 127–150.
- Croskey, C. L., et al., 1992. The millimeter wave atmospheric sounder (MAS): a shuttle-based remote sensing experiment. *IEEE Transactions on Microwave Theory and Techniques*, **40**(6), 1090.
- Drummond, J. R., Houghton, J. T., Peskett, G. D., Rodgers, C. D., Wale, M. J., Whitney, J. G., and Williamson, E. J., 1980. The stratospheric and mesospheric sounder on nimbus 7.

- Philosophical Transactions of the Royal Society of London*, **296**(1418), 219–241.
- Fischer, H., et al., 2008. MIPAS: an instrument for atmospheric and climate research. *Atmospheric Chemistry and Physics*, **8**, 2151–2188.
- Gille, J. C., and Russell, J. M., III, 1984. The limb infrared monitor of the stratosphere: experiment description, performance and results. *Journal of Geophysical Research*, **89**, 5125–5140.
- Gille, J., et al., 2008. High resolution dynamics limb sounder (HIRDLS): experiment overview, recovery and validation of initial temperature data. *Journal of Geophysical Research*, **113**, D16S43, doi:10.1029/2007JD008824.
- Gunson, M. R., et al., 1996. The atmospheric trace molecule spectroscopy (ATMOS) experiment: deployment on the ATLAS space shuttle missions. *Geophysical Research Letters*, **23**, 2333–2336.
- Hays, P. B., Abreu, V. J., Dobbs, M. E., Gell, D. A., Grassi, H. J., and Skinner, W. R., 1993. The high-resolution doppler imager on the upper atmosphere research satellite. *Journal of Geophysical Research*, **98**, 10, 713–10, 723.
- Livesey, N. J., and Read, W. G., 2000. Direct retrieval of line-of-sight atmospheric structure from limb sounding observations. *Geophysical Research Letters*, **27**, 891–894.
- Llewellyn, E. J., et al., 2003. First results from the OSIRIS instrument on-board Odin. *Sodankyla Geophysical Observatory Publications*, **92**, 1–47.
- Lucke, R. L., et al., 1999. The polar ozone and aerosol measurement (POAM) III instrument and early validation report. *Journal of Geophysical Research*, **104**, 18785–18799.
- McCleese, D. J., et al., 2007. Mars climate sounder: an investigation of the thermal and water vapor structure, dust and condensate distributions in the atmosphere, and energy balance of the polar regions. *Journal of Geophysical Research*, **112**, E05S06, doi:10.1029/1006JE002790.
- McCormick, M. P., Zawodny, J. M., Velga, R. E., Larsen, J. C., and Wang, P. H., 1989. An overview of SAGE I and II ozone measurements. *Planetary and Space Science*, **37**, 1567–1586.
- Murtagh, D., et al., 2002. An overview of the Odin atmospheric mission. *Canadian Journal of Physics*, **80**, 309–319.
- Ozeki, H., et al., 2001. Development of superconducting submillimeter-wave limb emission sounder (JEM/SMILES) aboard the International Space Station. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Bellingham: SPIE, Vol. 4540, pp. 209–220.
- Roche, A. E., Kumer, J. B., Mergenthaler, J. L., Ely, H. A., Uplinger, W. H., Potter, J. F., James, T. C., and Sterritt, L. W., 1993. The cryogenic limb array etalon spectrometer (CLAES) on UARS: experiment description and performance. *Journal of Geophysical Research*, **98**, 10, 763–10, 775.
- Rodgers, C. D., 2000. *Inverse Methods for Atmospheric Science, Theory and Practice*. Singapore: World Scientific.
- Russell, J. M., III, et al., 1993. The halogen occultation experiment. *Journal of Geophysical Research*, **93**, 10, 777–10, 798.
- Russell, J. M., III, Mlynczak, M. G., Gordley, L. L., Tansock, J., and Esplin, R. 1999. An overview of the SABER experiment and preliminary calibration results. Proc. SPIE, 3756, doi:10.1117/12.366382, 277–288.
- Shepherd, G. G., et al., 1993. WINDII, the wind imaging interferometer on the upper atmosphere research satellite. *Journal of Geophysical Research*, **98**, 10, 725–10, 750.
- Taylor, F. W., et al., 1993. Remote sensing of atmospheric structure and composition by pressure modulation radiometry from space: the ISAMS experiment on UARS. *Journal of Geophysical Research*, **98**, 10, 799–10, 814.
- Waters, J. W., et al., 2006. The earth observing system microwave limb sounder (EOS MLS) on the Aura satellite. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1075–1092.

Cross-references

[GPS, Occultation Systems](#)
[Stratospheric Ozone](#)
[Trace Gases, Stratosphere, and Mesosphere](#)

M

MADDEN-JULIAN OSCILLATION (MJO)

Baijun Tian and Duane Waliser
Jet Propulsion Laboratory, California Institute of
Technology, Pasadena, CA, USA

Synonyms

30–60 day oscillation; 40–50 day oscillation;
Intraseasonal oscillation (ISO); Intraseasonal variability
(ISV)

Definition

The MJO is a planetary-scale quasiperiodic oscillation of atmospheric wind and convective cloudiness anomalies that moves slowly eastward along the equator mainly over the tropical Indian and Pacific Oceans with a timescale on the order of 30–60 days.

Introduction

In 1971, Roland Madden and Paul Julian stumbled across a 40–50 day oscillation when analyzing the zonal (east–west) wind data from rawinsondes at Kanton Island (3 °S, 172 °W) over the equatorial western Pacific. Until the early 1980s, little attention was paid to this oscillation, which later became known as the MJO. Since the 1982–1983 El Niño event, low-frequency variations in the tropics, both on intra-annual (less than a year) and interannual (more than a year) timescales, have received much more attention, and the number of MJO-related publications grew rapidly. The MJO turned out to be the dominant form of the intraseasonal (30–90 day) variability in the tropical atmosphere and has many important influences on the global weather and climate system.

The MJO is a naturally occurring mode of variability of the tropical ocean–atmosphere system. It is characterized by an eastward propagation of large regions of both

enhanced and suppressed tropical convection, cloudiness and rainfall near the equator mainly over the tropical Indian and Pacific Oceans, and associated large-scale atmospheric circulation (wind) anomalies over the whole globe. The anomalous cloudiness or rainfall usually first emerges over the equatorial western Indian Ocean and intensifies and remains evident as it propagates eastward over the warm ocean waters of the equatorial eastern Indian Ocean and western Pacific, the so-called Indo-Pacific warm pool. This pattern of anomalous cloudiness and rainfall then generally weakens and disappears as it moves over the cooler ocean waters of the equatorial eastern Pacific, the so-called equatorial cold tongue. Along with this eastward-propagating pattern of equatorial cloudiness and rainfall anomalies, there also exist eastward moving distinct baroclinic patterns of lower- and upper-level atmospheric circulation anomalies in the tropics and subtropics. The circulation anomalies extend around the globe and are not confined to the eastern hemisphere as opposed to the cloudiness and rainfall anomalies. When the MJO moves eastward, it modulates the background cloud, rainfall, and circulation in the tropical Indian and Pacific Oceans on timescales shorter than a season but longer than a couple of weeks. The length of a typical MJO cycle is approximately 30–60 days but normally 40–50 days. Thus, the MJO is also known as the 30–60 day oscillation, 40–50 day oscillation, intraseasonal oscillation (ISO), or intraseasonal variability (ISV) after its typical timescale. A complete MJO cycle can be divided into two distinct phases according to the intensity of its convective activity and rainfall: convectively active (enhanced) phase or wet phase and convectively inactive (suppressed) phase or dry phase. The wet phase of the MJO cycle is characterized by enhanced tropical convection and large moist convective storms with higher cloud-top heights, more cloud cover, and heavier rainfall (thus more atmospheric latent heating)

than average. In contrast, the dry phase of the MJO cycle is typified by dry and clear conditions with lower cloud-top heights, less cloud cover and rainfall (thus less atmospheric latent heating) than normal. The MJO appears to be predictable with lead times of 2–3 weeks. This can help bridge the gap in environmental forecast skill between that of weather (lead times up to a few days) and that of seasonal-to-interannual climate predictions (lead times from a few months to a few years).

Satellite remote sensing data have played an important role in the MJO studies during the last three decades because of the high spatial (a few kilometers) and temporal (3 h or daily) resolutions and global coverage of the satellite data especially over the tropical oceans where the rawinsondes are sparse and the global reanalyses have had large uncertainties. These satellite-based studies have significantly advanced our knowledge in the MJO description and mechanism as well as its global impacts. They have also led to considerable improvement in our numerical modeling capability and theoretical understanding of the MJO. This entry briefly reviews the central role of satellite remote sensing data in studying the description, mechanisms, and global impacts of the MJO.

Description

The MJO and its eastward-propagating convective feature are most active during the Northern Hemisphere (boreal) winter season (November–April) when the Indo-Pacific warm pool is centered near the equator. During the Northern Hemisphere (boreal) summer season (May–October), the change in the large-scale wind patterns associated with the Asian summer monsoon results in the large-scale convective disturbances propagating northeastward, from the equatorial Indian Ocean into Southeast Asia. The discussions in this entry mainly focus on the boreal winter MJO events although many aspects of these discussions can be equally applied to the boreal summer MJO events.

The MJO can be detectable in several important atmospheric and oceanic parameters, such as atmospheric cloudiness, atmospheric wind speed and direction, atmospheric temperature, atmospheric moisture, surface pressure, surface rainfall, [sea surface temperature \(SST\)](#), and surface heat and freshwater fluxes. However, the fundamental quantities related to the MJO are large-scale organized convection, cloudiness, rainfall, and tropospheric winds. Thus, our discussion of the MJO description mainly centers on how the remote sensing data help us understand the characteristics of large-scale convective cloudiness, rainfall, and tropospheric winds associated with the MJO.

Outgoing longwave radiation (OLR) and infrared window radiance (or brightness temperature) provide broadband and narrowband measures of the total flux of longwave radiation lost to space at the top of the atmosphere. Deep convective clouds in the tropics have cold cloud tops and therefore have low values of OLR and brightness temperature. Typically, an OLR value of less

than 200 W m^{-2} or a brightness temperature value of less than 220 K indicates the presence of deep convection in the tropics. Because of this simple property of OLR and brightness temperature, they have been widely used as a proxy for deep convection over the tropics, where the background longwave radiation from low clouds or surface is much higher. During the wet phase of the MJO, both OLR and window radiance can be significantly reduced due to higher and colder convective cloud tops. On the other hand, during the dry phase of the MJO, both OLR and window radiance can be significantly enhanced because of clear skies and lower and warmer cloud tops. Thus, OLR and infrared window radiance are good indicators of the convective intensity of the MJO. Over the tropical warm ocean waters where MJO convection is active, satellites are the only way to observe convective cloudiness with large spatial coverage. As a result, the satellite remote sensing data have played a central role in studying the general spatial and temporal structure and eastward-propagating features of the MJO convection and cloudiness during the past three decades.

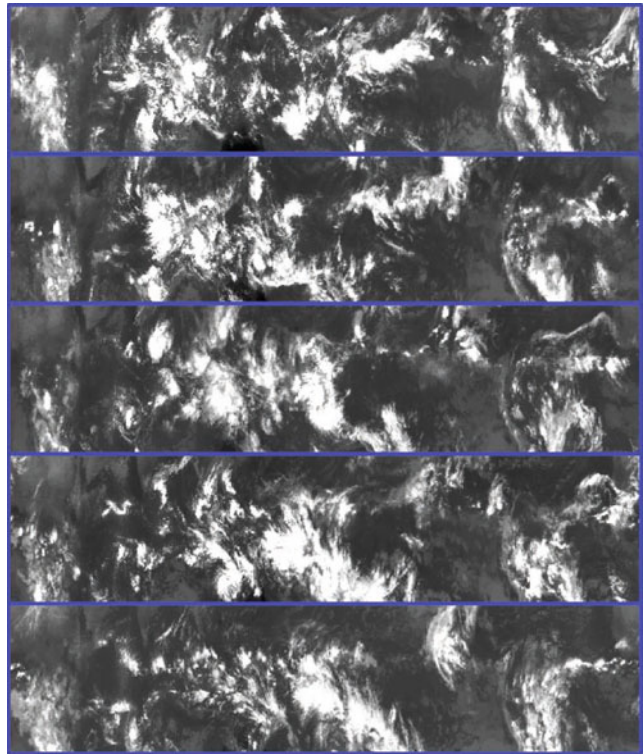
Based on limited rawinsonde and surface station data, Madden and Julian speculated that the MJO is characterized by slowly eastward-propagating, large-scale oscillations in the tropical convective cloudiness over the equatorial Indian Ocean and western Pacific as the result of an eastward movement of large-scale atmospheric circulation cells oriented in the equatorial zonal plane. Evidence of such eastward-propagating clouds in satellite data was first presented by Arnold Gruber in 1974 and Abraham Zangvil in 1975 who both found large-scale eastward-propagating features near 40–50 days at the equator in the cloud brightness data obtained from the Environmental Science Services Administration (ESSA) satellites (ESSA 3 and 5). However, no further evidence was found until the early 1980s when NOAA OLR data and wind analyses from US National Meteorological Center (NMC) became available. These Advanced Very High Resolution Radiometer (AVHRR) OLR data started from the mid-1970s and were mainly from a series of polar-orbiting satellites, such as the scanning radiometer (SR) series and the Television Infrared Observation Satellite–Next Generation (TIROS-N) series. These OLR data have a twice-daily resolution and a good global coverage each day. By the early 1980s, almost 10 years of daily AVHRR OLR data were archived and available to the research community. In the early to mid-1980s, a series of observational papers on the MJO (at the time still referred to as the 40–50 day or 30–60 day oscillation) using the AVHRR OLR and NMC wind analyses appeared. These studies clearly demonstrated the existence of the slowly eastward propagation of the tropical cloudiness at the intraseasonal timescale and documented many detailed and important convective cloudiness and circulation features of the MJO. These papers also helped to bring the MJO to the attention of the scientific community. The NOAA polar orbiter satellites have been operating almost continuously over the

past 30 years. As a result, the NOAA OLR data have a relatively long record (over 30 years) and have been and are still extensively employed for studying the MJO.

Figure 1 shows the infrared brightness for a MJO event from December 7 to 26, 1987, with each panel separated by 5 days. Each map covers the all longitudes (0–360°) between 20°S and 20°N. Bright white areas indicate cold high clouds, and dark regions indicate cloud-free or warm low cloud conditions. The slow eastward propagation of cold high clouds associated with the MJO is evident. The cold high clouds first form over the western equatorial Indian Ocean on December 7, 1987 and, over the course of the following ~20 days, then intensify and propagate eastward across the equatorial Indian Ocean and the Maritime Continent to the equatorial western Pacific Ocean.

In addition to the large-scale eastward-propagating pattern of MJO convective cloudiness, there exist many fine-scale structures within the convective cloudiness. The high spatial (100 km) and temporal (daily) resolution AVHRR OLR data from NOAA polar orbiters and much higher spatial (a few kilometers) and temporal (3 h) resolution window-channel infrared data from the geostationary satellites, such as Geostationary Meteorological Satellite (GMS) from Japan, are particularly useful in investigating the fine structure of the MJO convective cloudiness due to their high spatial and temporal resolutions. For example, the OLR data indicated many short-period, synoptic-scale convective systems within the planetary-scale 30–60 day fluctuations. Along the equator, these active convective systems move eastward with a phase speed of 10–15 m s⁻¹ and have a horizontal spatial scale of several thousand kilometers and a timescale of less than 10 days. These synoptic-scale, eastward-propagating convective systems within the MJO envelope are referred to as super cloud clusters or more recently convectively coupled Kelvin waves. The GMS infrared window radiance data have revealed that a super cloud cluster consists of many fine-scale cloud clusters. These fine-scale cloud clusters typically propagate westward along the equator with a lifetime of about 1–2 days. Although each cloud cluster moves westward, a super cloud cluster moves eastward due to the successive formation of a new cloud cluster east of the mature-stage cloud cluster. This suggested the existence of a hierarchy of convective activity within the MJO that is still an outstanding avenue of research of today.

In the Tropics, surface rainfall is closely related to convective cloudiness and thus is another key quantity of interest to characterize the MJO. The tropical rainfall can be estimated from the satellite-observed infrared and microwave radiances. The surface rainfall can first be indirectly derived from infrared window radiance and OLR which are very sensitive to cloud-top temperatures that are indirectly tied to surface rainfall. The microwave radiances are very sensitive to the hydrometeors that directly result in surface precipitation and thus can be used more directly to retrieve surface precipitation. The microwave-based rainfall retrievals can be divided into passive and



Madden-Julian Oscillation (MJO), Figure 1 Infrared satellite observations for the global tropics (20°N–20°S) for (from the top) 03 GMT December 7, 12, 17, 22, and 26, 1987. Bright white areas indicate high clouds and deep convection, and dark regions indicate cloud-free conditions (Based on Global Cloud Imagery (GCI) data courtesy of M. Salby, University of Colorado).

active microwave (radar) retrievals. The passive retrievals can further be divided into the microwave emission-based (sensitive to cloud liquid water) and the microwave scattering-based (sensitive to ice particles and large water drops) rainfall retrievals. Starting from the 1990s, several global rainfall data have been generated from satellite-observed infrared and microwave data and were instrumental in studying the MJO during the last two decades. For example, daily, global oceanic rainfall data retrieved based on microwave emission from the Microwave Sounding Unit (MSU) on the NOAA TIROS-N satellites was first used to study the MJO convective feature in the 1990s. During the late 1990s, the NOAA Climate Prediction Center (CPC) generated a global rainfall data set, referred to CPC Merged Analysis of Precipitation (CMAP), through the merged analysis of precipitation from several sources, such as gauges, satellites, and numerical model outputs. The satellite rainfall estimates for the CMAP includes the window infrared-based rainfall estimate from NOAA geostationary satellites (e.g., Geostationary Operational Environmental Satellites, GOES), the OLR-based rainfall estimate from the NOAA polar-orbiting satellites, the microwave emission-based rainfall estimate from the MSU on the NOAA polar-orbiting

satellites and the Special Sensor Microwave/Imager (SSM/I) on the Defense Meteorological Satellite Program (DMSP) satellites, and the microwave scattering-based rainfall estimate from SSM/I. Similar merged global rainfall data were also produced by the Global Precipitation Climatology Project (GPCP) based on similar inputs. However, some subtle differences exist between the CMAP and GPCP precipitation data sets, such as diurnal cycle adjustment and atoll precipitation adjustments. Both CMAP and GPCP rainfall data with pentad (5 day) resolution have been extensively used to study the convective features of the MJO and identify MJO events. Launched in 1997, the Tropical Rainfall Measurement Mission (TRMM) satellite provided the first spaceborne precipitation radar (PR) (active microwave) to monitor global rainfall from space in addition to the passive TRMM Microwave Imager (TMI) instrument and the visible and infrared scanner (VIRS) instrument. The TRMM PR and TMI data have also been used to study the MJO. However, their spatial and temporal sampling is rather coarse. To alleviate these sampling deficiencies of the TRMM PR and TMI, the TRMM Multisatellite Precipitation Analysis (TMPA) project provides a calibration-based sequential scheme for combining precipitation estimates from multiple satellites, as well as gauge analyses where feasible, at fine scales ($0.25^\circ \times 0.25^\circ$ and 3 hourly). The input satellite data for the TMPA are mainly from two sources: (1) passive microwave-based precipitation from the TMI on TRMM, the SSM/I on DMSP satellites, the Advanced Microwave Scanning Radiometer–Earth Observing System (AMSR-E) on Aqua, and the Advanced Microwave Sounding Unit-B (AMSU-B) on the NOAA polar-orbiting satellites; (2) window infrared-based precipitation data collected by the international constellation of geostationary satellites. This TMPA data set, also known as the TRMM 3B42, has relatively better retrieval accuracy and sampling at fine spatial and temporal scales. It has been used extensively for the recent MJO studies.

In association with the eastward-propagating equatorial convective cloud and rainfall system are strong variations in lower- and upper-level large-scale atmospheric wind fields along the equator and in the subtropics. Unlike the convective cloudiness that is mostly confined over the equatorial Indian and western Pacific Oceans, the large-scale wind anomalies of the MJO extend globally along the equator and into the subtropics. For example, along the equator, low-level zonal winds converge into the convective center, while upper-level zonal winds diverge away from the convective center. These lower- and upper-level zonal winds are interconnected through ascending (upward vertical movement) moist air within the convective center and descending dry air outside the convective center. These large-scale zonal winds propagate eastward together with the convective cloudiness along the equator and can reach into the western hemisphere (eastern Pacific, Atlantic, and Africa). In addition to these zonal winds along the equator are large-scale gyre circulations extending into the subtropics in both the lower

and upper troposphere that are tied to the eastward-propagating convective cloudiness and zonal winds along the equator. For example, in the lower troposphere, a subtropical cyclonic couplet (counterclockwise in the Northern Hemisphere and clockwise in the Southern Hemisphere) flanks or lies to the west of the MJO convective region, while a subtropical anticyclonic couplet (clockwise in the Northern Hemisphere and counterclockwise in the Southern Hemisphere) lies to the east of the MJO convective region. On the other hand, in the upper troposphere, a subtropical anticyclonic couplet flanks or lies to the west of the MJO convective region, while a subtropical cyclonic couplet lies to the east of the MJO convective region due to the baroclinic nature of the tropical large-scale wind fields. The near equatorial large-scale zonal wind anomalies are a Kelvin wave response to the MJO convective heating, while the off equatorial meridional wind anomalies are a Rossby wave response to the MJO convective heating. The Kelvin wave is named after Lord Kelvin, who studied water waves along a vertical side boundary under rotation conditions. In this case, the equator, where the vertical component of the Earth's rotation vector changes sign, serves the vertical side boundary. The Rossby wave is due to the latitudinal variation of the vertical component of the Earth's rotation and is named after C. G. Rossby, who was the first to clearly isolate the so-called Rossby wave dynamics (a balance between inertia and rotation). These large-scale patterns of convective cloudiness and wind fields are components of what are collectively referred to as equatorial waves or convectively coupled equatorial waves. It is the upper-level circulation features of these waves that allow the convective signatures of the MJO over the Indo-Pacific warm pool to influence weather "downstream" over the eastern Pacific and Atlantic (e.g., hurricanes and tropical cyclones) as well as the midlatitudes (e.g., precipitation extremes along the US west coasts).

In terms of satellite observations, tropospheric winds are difficult to observe directly except via cloud tracking, such as the cloud-drift winds derived from the Multi-angle Imaging SpectroRadiometer (MISR) and NOAA geostationary satellites. However, the surface winds can be measured directly using the spaceborne radar scatterometers, such as the SeaWinds instrument on NASA's Quick Scatterometer (QuikSCAT) satellite. The importance of the cloud-drift winds and QuikSCAT surface winds for the MJO study has been recognized but still in the early stages of exploration. The overall large-scale dynamic structure of the MJO, especially in the upper levels, is still mainly derived from the global reanalyses or radiosondes at the moment.

Mechanisms

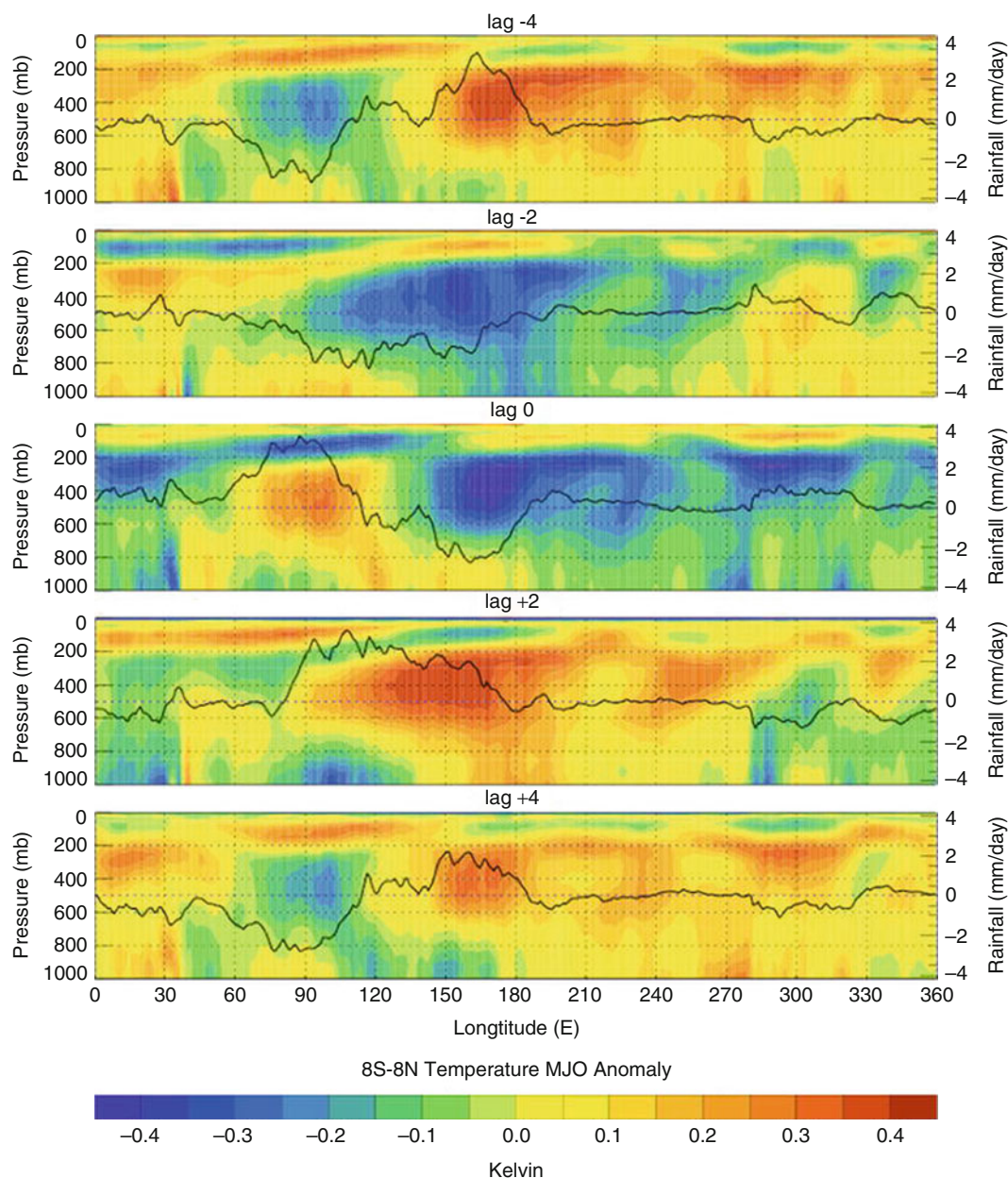
To understand the mechanisms responsible for the initiation and maintenance of the MJO, it is important to quantify the evolution of the thermodynamic environment and surface conditions associated with the MJO. In particular,

documenting the three-dimensional temperature and moisture structure of the MJO is crucial in advancing our theoretical understanding of the MJO. The availability of satellite-based temperature and moisture soundings makes these studies possible. For example, tropical mean tropospheric temperature, lower-troposphere (surface–300 mb) temperature, and upper-troposphere (500–100 mb) temperature were derived from the MSU channels in the early 1990s. These data were used to study the relationship between MJO convection and temperature anomalies in the 1990s. It was found that when the MJO is amplifying (e.g., over eastern Indian Ocean), convective heating anomalies are positively correlated to temperature anomalies. This implies production of eddy available potential energy (EAPE), which can in turn be used to drive atmospheric motion and sustain the MJO. When the MJO is decaying (e.g., over the eastern Pacific or east of the Date Line), temperature anomalies are nearly in quadrature with convective heating anomalies. As a result, their correlation and production of EAPE are small which is no longer an energy source for the MJO. For water vapor, the TIROS Operational Vertical Sounder (TOVS) provided the water vapor fields at five different levels in the troposphere in the 1990s. These data were used to study the three-dimensional structure and evolution of water vapor over the life cycle of the MJO in the early 2000s. The composite evolution of moisture shows markedly different vertical structures as a function of longitude. There is a clear westward tilt with the height of the moisture maximum associated with the MJO propagating eastward across the Indian Ocean. These disturbances evolve into nearly vertically uniform moist anomalies as they reach the western Pacific. Near-surface (below 850 mb) positive water vapor anomalies were observed to lead the convection anomaly by 5 days over the Indian Ocean and western Pacific. Upper-level positive water vapor anomalies were observed to lag the peak in the convection anomaly by 5–10 days, as the upper troposphere is moistened following intense convection. In the eastern Pacific, the moisture variations then become confined to the lower levels (below 700 mb), with upper-level water vapor nearly out of phase.

While the MSU and TOVS provided useful initial insights into the three-dimensional temperature and moisture structure of the MJO, their vertical resolution was too low to describe the detailed vertical structure, especially near the tropopause and boundary layer. Recently, global atmospheric moisture and temperature profiles with a much higher vertical resolution were produced by the Atmospheric Infrared Sounder (AIRS)/Advanced Microwave Sounding Unit on the NASA Aqua satellite. The AIRS data provide an unprecedented opportunity to document the vertical moist thermodynamic structure and spatial–temporal evolution of the MJO. Figure 2 presents the pressure–longitude cross sections of the temperature anomaly and its relationship to the convective rainfall anomaly for a composite MJO cycle (from –20 day to +20 day, separated by 10 days each). In the Indo-Pacific

warm pool, the temperature anomaly exhibits a trimodal vertical structure: a warm anomaly in the free troposphere (800–250 hPa) and a cold anomaly near the tropopause (above 250 hPa) and in the lower troposphere (below 800 hPa) for the wet phase and vice versa for the dry phase. The moisture anomaly also shows markedly different vertical structures as a function of longitude and the strength of the convection anomaly. Most significantly, the AIRS data demonstrate that, over the Indian Ocean and western Pacific, enhanced convection and precipitation is generally preceded in both time and space by a low-level warm and moist anomaly and followed by a low-level cold and dry anomaly. This zonal asymmetry in the low-level moisture and temperature anomaly provides a favorable moist thermodynamic condition for the eastward propagation of the MJO. Furthermore, the comparison between the AIRS observations and the National Center for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) and NCEP/Department of Energy (DOE) reanalyses revealed the poor representation of the low-level moisture and temperature structure associated with the MJO in these reanalyses particularly over the equatorial Indian and Pacific Oceans, where there are very few conventional data to constrain the reanalyses. Despite their reliance on (imperfect) numerical models, these reanalyses have been widely used as “observations” to validate MJO theories and model simulations.

The ocean surface fluxes of heat (solar and infrared radiation, latent and sensible heat), mass (rainfall and water vapor flux), and momentum (wind) are considered important for the initiation and eastward propagation of the MJO by some researchers. Understanding the manner and degree the atmospheric components of the MJO are coupled to ocean surface fluxes is vital to understand the MJO dynamics and establish MJO theories. This coupling includes how the atmospheric components of the MJO influence the ocean surface heat, mass, and momentum fluxes and how the ocean surface provides the needed heat and moisture sources for the MJO convection. Surface heat fluxes have typically been very difficult to measure remotely from space although satellites offer the only viable way to estimate these quantities with regular temporal and spatial samplings over the tropical oceans. Apart from precipitation discussed above, satellite measurements of clouds and water vapor have been used with atmospheric radiative transfer models to provide estimates of solar and infrared radiation fluxes at the ocean surface. Moreover, satellite estimates of ocean surface winds (e.g., QuikSCAT and SSM/I discussed above) have been used in conjunction with water vapor measurements from satellites to construct estimates of latent heat (or evaporative) flux from the ocean. These types of observations, in conjunction with satellite SST retrievals (e.g., Advanced Very High Resolution Radiometer (AVHRR) or TMI), have been used to study how the MJO convection interacts with the ocean surface and explore the degree the ocean and atmosphere are coupled at intraseasonal and other

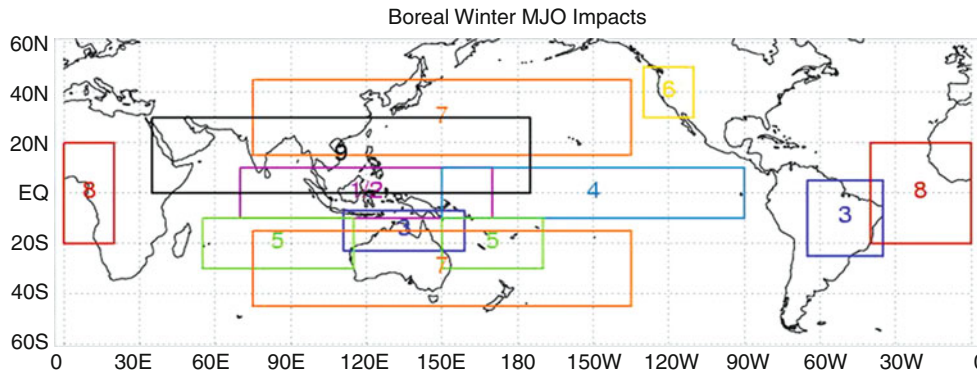


Madden-Julian Oscillation (MJO), Figure 2 Pressure-longitude cross sections of equatorial mean (averaged from 8°S to 8°N) temperature anomaly for a composite MJO cycle based on AIRS data from 2002 to 2005. The color *red* denotes warm anomalies, while the color *blue* indicates cold anomalies. The superimposed *solid black line* denotes rainfall anomaly from TRMM (Reproduced from Tian et al., 2006, their Figure 3, by permission of American Meteorological Society).

timescales. Satellite-based ocean surface wind speed observations (e.g., QuikSCAT and SSM/I discussed above) have been used to derive momentum fluxes (e.g., wind stress) at the ocean surface. These observations have been critical in documenting and understanding the role of the MJO in influencing the development and evolution of El Niño and Southern Oscillation (ENSO) events. ENSO is the most important interannual variability in the coupled tropical atmosphere ocean system with a dominant

timescale of 2–7 years and has significant impacts on the global weather and climate.

During the dry phase of the MJO, suppressed convection is associated with decreased cloud cover and increased surface insolation and anomalous surface easterlies. These anomalous surface easterlies act to decrease the surface wind speed because the background surface winds are weak westerlies in the equatorial Indian and western Pacific Oceans, hence decreasing the surface latent heat flux



Madden-Julian Oscillation (MJO), Figure 3 Geographical regions of a number of impacts of the boreal winter MJO on the global climate system. (1) Influencing tropical weather, alternative periods of wetter/drier conditions in the tropical Indian Ocean and western Pacific. (2) Modulating the diurnal cycle of tropical deep convection and rainfall in the tropical Indian Ocean and western Pacific. (3) Modulating the onsets and breaks of the Australian and South American monsoon systems. (4) Influencing ENSO cycle over the equatorial central and eastern Pacific. (5) Impacting tropical cyclone genesis over the tropical Indian Ocean and western Pacific. (6) Influencing the development of heavy rainfall events over US west coast. (7) Changing the subtropical total-column ozone over the eastern Hemisphere and Pacific Ocean. (8) Affecting the aerosol and air pollution over the equatorial Indian and western Pacific Oceans as well as the tropical Africa and Atlantic Ocean. (9) Influencing the ocean surface Chl across the tropical ocean coasts.

(or evaporation). Increased surface shortwave radiation and reduced surface evaporation contribute to the warming of SST for the dry phase. During the subsequent wet phase of the MJO, enhanced convection is associated with increased cloud cover and decreased surface insolation. As a result, the SST warming trend is arrested and a cooling trend is initiated. Subsequently, the continued cooling of the upper ocean is accelerated by increased westerly surface winds leading to enhanced surface evaporation and increased entrainment of cold water from below the thermocline. Then the wet phase is followed by another dry phase when SST warming occurs. Therefore, over the Indian Ocean and western Pacific, the enhanced convection is usually led by a warm SST anomaly to the east due to enhanced insolation and decreased evaporation and followed by a cold SST anomaly to the west due to decreased insolation and enhanced evaporation. When the convective anomaly approaches the Date Line, the surface evaporation anomaly and surface solar radiation anomaly tend to cancel each other. Thus, the SST anomaly is rather small over the eastern Pacific, so does the convective anomaly. This convection–SST phase relationship leads many scientists to believe that the MJO is a coupled mode of the tropical ocean–atmosphere system.

Impacts

During the past three decades, the MJO has been shown to have important influences on various weather and climate phenomena over the globe at many timescales, such as the diurnal cycle, tropical weather, monsoon onsets and breaks, ENSO, tropical hurricanes and cyclones, extreme precipitation events, extratropical and high-latitude circulation, and weather patterns. Given evidence that the MJO is predictable with lead times of 2–3 weeks, the strong modulation of the global climate system by the MJO implies that many

other components of the global climate system may be predictable with similar lead times. Some examples of the MJO impacts are listed and described below, and the regions of impacts are illustrated in Figure 3.

First, the MJO significantly impacts the tropical synoptic weather, such as alternative periods of wet and dry conditions, especially over the tropical Indian Ocean and west Pacific, through its influences of tropical rainfall and cloudiness (Figure 3). During the wet phase of the MJO, the tropical atmosphere is very moist and cloudy with heavy rainfall. In contrast, during the dry phase of the MJO, the tropical atmosphere experiences dry and clear conditions with plenty of sunshine.

Second, the MJO can influence the diurnal cycle of tropical deep convection and rainfall through its effect on the background state over the equatorial Indian and western Pacific oceans (Figure 3). The diurnal cycle is enhanced over both land and water during the convectively active phase of the MJO, while it is reduced during the convectively suppressed phase of the MJO. However, the diurnal phase is not significantly affected by the MJO.

Third, the MJO can substantially modulate the intensity of monsoon systems around the globe, such as the Australian and South American monsoons for boreal winter and the Asian and North American monsoons for boreal summer (Figure 3). The wet phase of the MJO can affect both the onset timing and intensity of the monsoon, while the dry phase of the MJO can prematurely end a monsoon and also initiate breaks during already existing monsoons.

Fourth, there is evidence that the MJO influences the ENSO cycle (Figure 3). It was argued that the westerly wind bursts associated with the MJO over the equatorial western Pacific are an important trigger for an El Niño event. The MJO may not cause an El Niño event, but

can contribute to the speed of development, and perhaps the overall intensity of an ENSO cycle.

Fifth, the MJO is known to modulate tropical cyclone activity in the Indian Ocean, Pacific Ocean, Gulf of Mexico, and Atlantic Ocean by providing a large-scale environment that is favorable (unfavorable) for storm development (Figure 3). For example, westerly wind anomalies at the surface in and just behind the area of enhanced convection of the MJO may generate cyclonic (anticyclonic) rotation north (south) of the equator, respectively. At the same time, in the upper levels, anticyclonic (cyclonic) rotation develops along and just behind the area of enhanced convection resulting in a means to reduce vertical wind shear and increase upper-level divergence – both of which are favorable for tropical cyclone development and intensification. The strongest tropical cyclones tend to develop during the wet phase of the MJO. As the MJO progresses eastward, the favored region for tropical cyclone activity also shifts eastward from the Indian Ocean to the Pacific Ocean and eventually to the Atlantic Ocean.

Sixth, boreal winter extreme precipitation events along the US west coast are often connected with the pattern of tropical rainfall and circulation anomalies associated with the MJO (Figure 3). When the heavy tropical rainfall associated with the MJO is concentrated at the Maritime Continent, a strong blocking anticyclone is located in the Gulf of Alaska with a strong polar jet stream around its northern flank. During this time, the US west coast typically experiences a dry spell. When the enhanced tropical rainfall associated with the MJO shifts to the central Pacific and weakens, a deep low pressure system typically forms near the Pacific Northwest coast and can bring up to several days of heavy rain and possible flooding to the Pacific Northwest coast. These events are often referred to as “Pineapple Express” events, so named because a significant amount of deep tropical moisture traverses the Hawaiian Islands on its way toward western North America.

Although most quantities/processes/phenomena of interest, such as hurricanes, monsoons, and extratropical circulation and weather patterns, are not wholly described by satellite data, the satellite-based OLR or rainfall data were usually used to identify the MJO events and often to characterize aspects of the impacts (e.g., rainfall). Some studies totally depend on the available satellite data. For example, in the study of the MJO impact on the diurnal cycle of tropical deep convection, the International Satellite Cloud Climatology Project cloud product was employed to characterize the diurnal cycle, and the TRMM 3B42 precipitation product was used to identify the MJO events.

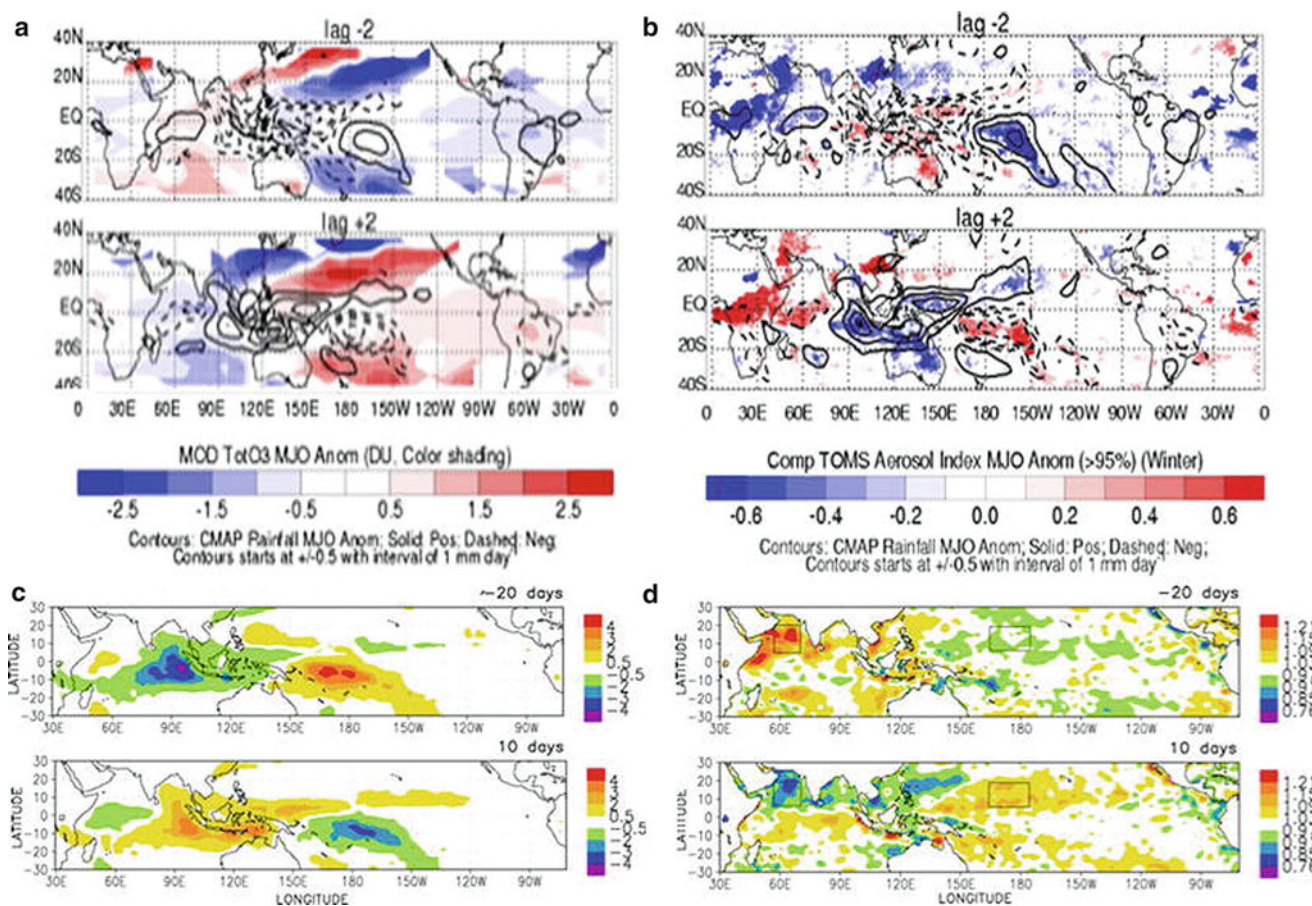
Recently, a number of studies have documented the MJO impacts on atmospheric composition, air quality, and biogeochemical cycle. This discovery has critically depended on the availability of satellite data. For example, the MJO

impact on tropical total-column ozone has been recently characterized using the satellite-observed tropical total-column ozone from the AIRS and Total Ozone Mapping Spectrometer (TOMS). It was found that tropical total ozone intraseasonal variations are large ($\sim \pm 10$ Dobson unit) and comparable to those in the annual and interannual timescales. These intraseasonal total ozone anomalies are mainly evident in the subtropics over the Pacific and eastern hemisphere, with a systematic relationship to the MJO convection and wave dynamics discussed earlier (Figures 3 and 4a). The subtropical negative ozone anomalies typically flank or lie to the west of the equatorial anomalous convection and are collocated with the subtropical upper-troposphere anticyclones generated by the equatorial anomalous convective heating. On the other hand, the subtropical positive ozone anomalies generally lie to the east of the equatorial anomalous convection and are collocated with the subtropical upper-troposphere cyclones generated by the equatorial anomalous convective heating. The subtropical ozone anomalies are anticorrelated with geopotential height anomalies near the tropopause and thus mainly associated with the ozone variability in the stratosphere rather than the troposphere.

Another example, the recent availability of multiple, global satellite aerosol products from TOMS, Moderate Resolution Imaging Spectroradiometer (MODIS), and Advanced Very High Resolution Radiometer (AVHRR) has made the investigation of the MJO modulation of aerosols possible. Large aerosol variations are found over the equatorial Indian and western Pacific Oceans where MJO convection is active, as well as the tropical Africa and Atlantic Ocean where MJO convection is weak, but the background aerosol level is high (Figures 3 and 4b). Although significant uncertainties still exist in the satellite aerosol retrievals, the satellite data indicate that the MJO and its associated cloudiness, rainfall, and circulation variability may systematically influence the aerosol variability.

The impacts of the MJO on the carbon monoxide (CO) abundances in the tropical tropopause layer (TTL) were also recently reported based on the Aura Microwave Limb Sounder (MLS) CO data. The effects of the eastward propagation of MJO on CO abundances in the TTL are evident. This indicates that the anomalous deep convection associated with the MJO can inject CO from the lower troposphere up to the TTL.

The availability of satellite-derived ocean surface chlorophyll (Chl) from Sea-viewing Wide Field-of-view Sensor (SeaWiFS) provided a means for the discovery of the MJO impacts on oceanic biology. It was found that the MJO produces systematic and significant variations in ocean surface Chl in a number of regions across the tropical Oceans, including the northern Indian Ocean, a broad expanse of the northwestern tropical Pacific Ocean, and a number of near-coastal areas in the far eastern Pacific Ocean (Figures 3 and 4c, d). This indicates a need to further investigate the MJO modulation of the biogeochemical cycle properties and higher levels of food chains.



Madden-Julian Oscillation (MJO), Figure 4 (a) Map of total-column ozone anomaly for a composite MJO cycle based on TOMS/SBUV data from 1980 to 2006. The color red denotes high ozone anomalies, while the color blue indicates low ozone anomalies. The superimposed solid black line denotes rainfall anomaly from CMAP (Reproduced from Tian et al., 2007, by permission of American Geophysical Union). (b) As (a) but for TOMS aerosol index anomaly (Reproduced from Tian et al., 2008, by permission of American Geophysical Union). (c) As (a) but for CMAP rainfall anomalies. (d) As (a) but for SeaWiFS ocean surface Chl anomaly (Reproduced from Waliser et al., 2005, by permission of American Geophysical Union).

Summary

The MJO is a large-scale quasiperiodic oscillation of tropical atmospheric circulation and convection anomalies that moves slowly eastward along the equator mainly over the tropical Indian and Pacific Oceans with a timescale on the order of 30–60 days. The MJO is the dominant form of the intraseasonal variability in the tropical atmosphere and has many important influences on the global weather and climate system. Since the 1970s, the satellite remote sensing data have played a fundamental role in advancing our knowledge in the MJO, particularly in terms of its description, theoretical mechanisms, and global impacts. First, the satellite data provided us the fundamental knowledge of the convective and dynamic features of the MJO. Second, the satellite data presented us the three-dimensional thermodynamic structure and the surface condition (e.g., SST and surface heat flux) evolution associated with the MJO that helped us to better understand the MJO and propose theoretical description. Third, the satellite data offered us the opportunity to discover

the global impacts of the MJO that have relevance to societal concerns such as extreme precipitation events, atmospheric composition, air quality, and biological markers in the ocean.

Acknowledgment

This research was carried out at Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Bibliography

- Gottschalk, J., Kousky, V., Higgins, W., and L'Heureux, M., 2008. Madden-Julian oscillation. <http://www.cpc.noaa.gov/products/precip/CWlink/MJO/mjo.shtml>
- Hendon, H. H., and Salby, M. L., 1994. The life-cycle of the Madden-Julian oscillation. *Journal of the Atmospheric Sciences*, **51**(15), 2225–2237.
- Lau, W. K. M., and Waliser, D. E. (eds.), 2005. *Intraseasonal Variability of the Atmosphere–Ocean Climate System*. Heidelberg: Springer. 474 pp.

- Madden, R. A., 2003. Tropical meteorology: intraseasonal oscillation (Madden-Julian oscillation). In Holton, J. R., et al. (eds.), *Encyclopedia of Atmospheric Sciences*. London: Academic, pp. 2334–2338.
- Madden, R. A., and Julian, P. R., 1971. Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific. *Journal of the Atmospheric Sciences*, **28**(7), 702–708.
- Madden, R. A., and Julian, P. R., 1972. Description of global-scale circulation cells in tropics with a 40–50 day period. *Journal of the Atmospheric Sciences*, **29**(6), 1109–1123.
- Madden, R. A., and Julian, P. R., 1994. Observations of the 40–50-day tropical oscillation: a review. *Monthly Weather Review*, **122**(5), 814–837.
- Tian, B., Waliser, D. E., Fetzer, E. J., Lambrigtsen, B. H., Yung, Y. L., and Wang, B., 2006. Vertical moist thermodynamic structure and spatial-temporal evolution of the MJO in AIRS observations. *Journal of the Atmospheric Sciences*, **63**(10), 2462–2485, doi:10.1175/JAS3782.1.
- Tian, B., Yung, Y. L., Waliser, D. E., Tyranowski, T., Kuai, L., Fetzer, E. J., and Irion, F. W., 2007. Intraseasonal variations of the tropical total ozone and their connection to the Madden-Julian oscillation. *Geophysical Research Letters*, **34**(8), L08704, doi:10.1029/2007GL029451.
- Tian, B., Waliser, D. E., Kahn, R. A., Li, Q. B., Yung, Y. L., Tyranowski, T., Geogdzhayev, I. V., Mishchenko, M. I., Torres, O., and Smirnov, A., 2008. Does the Madden-Julian oscillation influence aerosol variability? *Journal of Geophysical Research*, **113**(D12), D12215, doi:10.1029/2007JD009372.
- Waliser, D. E., 2006. Intraseasonal variability. In Wang, B. (ed.), *The Asian Monsoon*. New York: Springer/Praxis, pp. 203–257.
- Waliser, D. E., Murtugudde, R., Strutton, P., and Li, J. L., 2005. Subseasonal organization of ocean chlorophyll: prospects for prediction based on the Madden-Julian oscillation. *Geophysical Research Letters*, **32**(23), L23602, doi:10.1029/2005GL024300.
- Zhang, C. D., 2005. Madden-Julian oscillation. *Reviews of Geophysics*, **43**(2), RG2003, doi:10.1029/2004RG000158.

Cross-references

Aerosols
Radars

MAGNETIC FIELD

Nils Olsen

DTU Space, Technical University of Denmark, Lyngby, Denmark

Definition

Core field. The main part of the geomagnetic field, caused by dynamo action in the Earth's fluid outer core.

Crustal field. The magnetic field contribution caused by (permanent or remnant) magnetized material in the Earth's crust.

Internal sources. Magnetic field contributions caused by electrical currents or magnetized material in the Earth's interior.

External sources. Magnetic field contributions produced by electric currents in the ionosphere or magnetosphere.

Introduction

The Earth has a large and complicated magnetic field, the major part of which is produced by a self-sustaining dynamo operating in the fluid outer core. Magnetic field observations provide one of the few tools for remote sensing the Earth's deep interior, especially regarding the dynamics of the fluid flow at the top of the core. However, what is measured at or near the surface of the Earth is the superposition of the core field and fields caused by magnetized rocks in the Earth's crust, by electric currents flowing in the ionosphere, magnetosphere, and oceans, and by currents induced in the Earth by time-varying external fields. These sources have their specific characteristics in terms of spatial and temporal variations, and their proper separation, based on magnetic measurements, is a major challenge. Such a separation is a prerequisite for remote sensing by means of magnetic field observations.

Data sources

Prior to the satellite era, only near-surface (ground-based, marine, and airborne) magnetic observations were available for magnetic probing the Earth's interior. Presently about 150 geomagnetic observatories monitor the time changes of the field; their data are available through the World Data Center system (e.g., www.ngdc.noaa.gov/wdc, www.wdc.kugi.kyoto-u.ac.jp, www.wdc.bgs.ac.uk) and INTERMAGNET (www.intermagnet.org). However, the global distribution of the observatory network is very uneven, with large uncovered areas especially over the oceans.

While observatories monitor the field change at a given fixed location, regional surveys map the magnetic field at a given epoch (short period fluctuations of external origin have to be removed from the raw data; for this purpose, measurements of the field change at a nearby observatory or a temporary station are used). However, only parts of the globe are presently covered by airborne and marine surveys. The obtained data have been used for instance in the preparation of the World Digital Magnetic Anomaly Map (WDMAM, Korhonen et al. (2007), see also <http://projects.gtk.fi/WDMAM/>) and dedicated anomaly maps like EMAG2 (Maus et al., 2009).

True global coverage with magnetic field observations is only possible with satellites. Although spaceborne probing began more than 60 years ago with the launch of the *Sputnik 3* satellite in 1958, the data coverage that is necessary for global field modeling was first obtained by the *POGO* satellite series (Cain, 2007) which measured the magnetic field intensity between 1965 and 1972. The first high-precision vector measurements from space were taken by the Magsat satellite (Purucker, 2007) in 1979–1980. More recently, the launch of the satellites Ørsted in February 1999 (Olsen, 2007), CHAMP in July 2000 (Maus, 2007), and SAC-C in November 2000 opened revolutionary new possibilities for probing the Earth from space. In the near future, the *Swarm* satellite constellation mission (Friis-Christensen et al., 2006, 2009), comprising

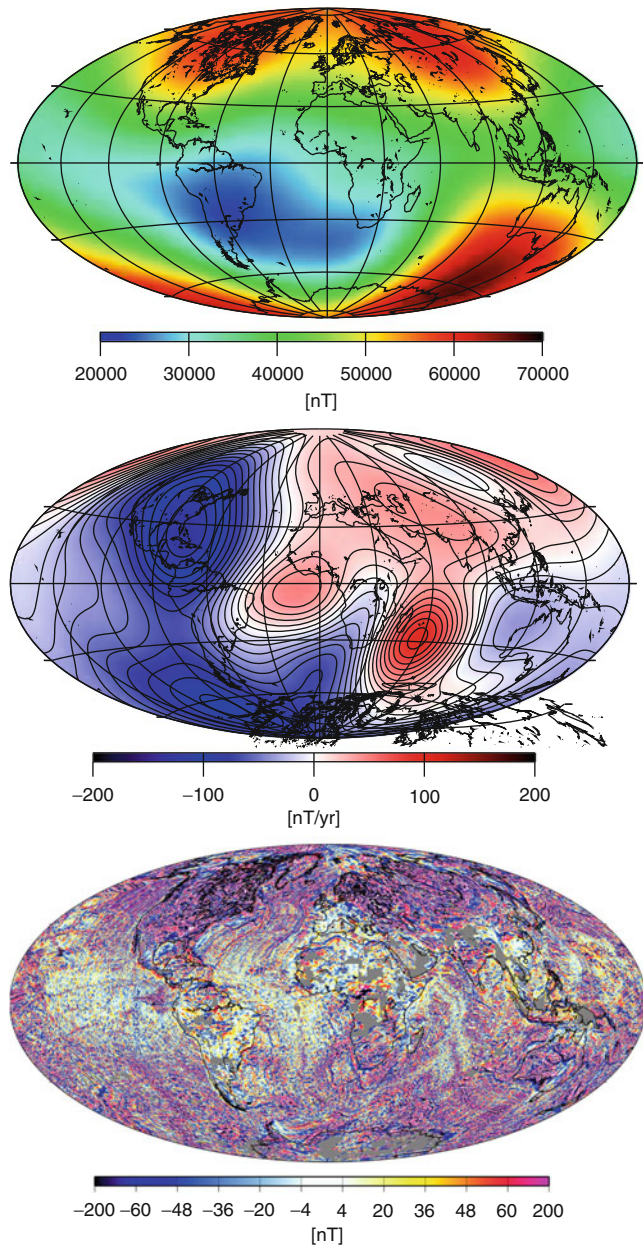
three satellites expected to be launched in late 2013, will provide even better possibilities for magnetic remote sensing the Earth's interior.

Characteristics of the geomagnetic field

The strength of the magnetic induction $\mathbf{B}$, in following for simplicity denoted as “magnetic field,” varies at Earth's surface between about 25,000 nT near the equator and about 65,000 nT near the poles ($1 \text{ nT} = 10^{-9} \text{ T}$, with $1 \text{ T} = 1 \text{ tesla} = 1 \text{ V/s}^{-1} \text{ m}^{-2}$). By far the largest part (95 % or more at Earth's surface) is due to dynamo action in the core (e.g., Roberts, 2007); magnetized material in the crust (e.g., Purucker and Whaler, 2007) accounts on average for only a few % of the total field but can locally reach magnitudes of several hundreds or even thousands of nT. Crustal magnetization consists of two parts: *Induced magnetization* is proportional, both in strength and direction, to the ambient field within which the rock is embedded. Were the core field to disappear, induced magnetization would vanish, too. Then, only the second type of magnetization, *remnant magnetization*, remains. As a general rule, remnant magnetization is weak in continental regions (where induced magnetization dominates), while both types of magnetizations are significant in oceanic areas.

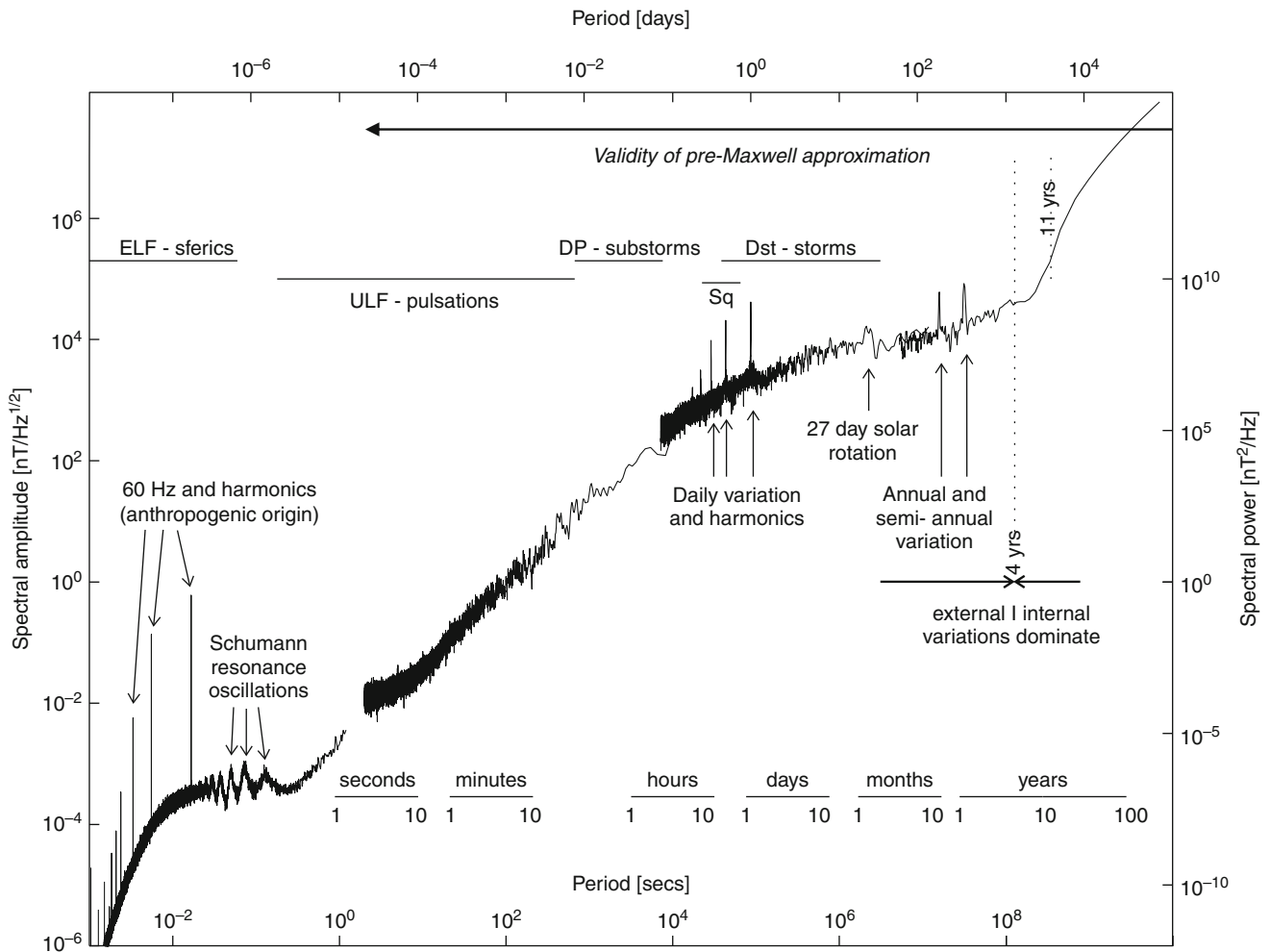
The top panel of Figure 1 shows the magnetic field intensity at Earth's surface in 2010.0, and the middle panel shows the yearly change in field intensity (also for epoch 2010.0), based on the International Geomagnetic Reference Field (IGRF) (Finlay et al., 2010). Core and crust contribute to all spatial scales of the field, but the core field dominates for horizontal scales larger than 2,800 km (corresponding to spherical harmonic degrees $n < 14$) while the crustal field dominates for horizontal scales smaller than 2,800 km ($n > 14$). The bottom panel of the figure shows the crustal anomaly field (scales smaller than 2,400 km) close to the surface (4 km above the geoid), based on the EMAG2 model of Maus et al. (2009). The crustal field is static (at least on timescales of centuries or shorter), but the core field undergoes a significant time change, known as secular variation. The temporal change shown in the middle panel of Figure 1 is therefore caused by processes in the Earth's core.

In addition to these internal sources, there are contributions from electric currents in the ionosphere (90–1,000 km altitude) and the magnetosphere (at distance larger than several Earth radii); these are called external sources. They are very dynamic, ranging from a few nT during geomagnetic quiet conditions to several hundreds or even thousands of nT during disturbed times, with especially large amplitudes at polar latitudes. Figure 2 shows a typical spectrum of the magnetic North component for a site at mid-latitudes. Time changes with periods longer than 4 years are dominated by core processes, while those at shorter periods, and especially signals with seasonal and daily periodicity, are caused by ionospheric and magnetospheric sources. See Schmucker (1985) for details on the



Magnetic Field, Figure 1 Earth's magnetic field intensity (*top*) and its time change (*middle*) in 2010, at Earth's surface. *Bottom*: intensity of magnetic field anomalies at 4 km altitude above the geoid.

external field variations ELF, sferics, ULF, pulsations, DP, Dst, and Sq. These short period changes, which are caused by external currents, produce secondary, induced, currents in the Earth's interior, the magnetic field of which adds to that of the primary, external, currents. The observed magnetic time variations are thus a superposition of primary (inducing) and secondary (induced) contributions. Analysis of these time changes



Magnetic Field, Figure 2 Amplitude of magnetic field variations (*north component*) at middle latitudes in dependence on period, determined using data from the sites Socorro/USA (1 ms to 1 s), Kakioka/Japan (2 s to 2 h), and Niemegk/Germany (longer than 2 h). Note that the pre-Maxwell approximation Equation 1 is only valid for periods $\gg 1$ s.

provides information on the electrical conductivity of the Earth's interior (e.g., Constable, 2007).

Although field changes of internal as well as external origin occur at all timescales, a common practice in separating them relies on their different temporal variations. Over the last 150 years, the axial dipole component of the Earth's magnetic field has decayed by nearly 10 %. This is ten times faster than the natural decay, in case the dynamo was switched off. The current decay rate is characteristic of magnetic reversals, which – as paleomagnetic data have shown – occur on average about once every half million years. Geographically, the recent dipole decay is largely due to changes in the field beneath the South Atlantic Ocean, connected to the growth of the South Atlantic anomaly. The core field and its temporal change (secular variation) directly reflect the fluid flow in the outermost core and provide a unique observational constraint on geodynamo theory. However, only the part of the core field that varies on timescales longer than, say, 1 year is

observable at the Earth's surface; shorter fluctuations are heavily attenuated due to the electrical conductivity of the mantle. Hence, variations with timescales longer than a few years are usually attributed to processes in the core, whereas those with periods shorter than 1 year are attributed to external field contributions. Yet interesting features occur at intermediate timescales. A serious limitation regarding the investigation of core processes at timescales of months to years is the effect of geomagnetic variations of external origin, since they contribute significantly on timescales up to that of the 11 year solar cycle.

Basic equations

Magnetic field investigation of core and crust is typically done in the quasi-static approximation, which requires that the timescales in consideration are longer ($\gg 1$ s) compared to the time required for light to pass the length scale of interest ($< a$ few thousand kilometers). In this so-called

pre-Maxwell approximation, displacement currents can be neglected, and the magnetic field $\mathbf{B}$ is given by

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \quad (1)$$

where $\mu_0 = 4\pi 10^{-7}$ Vs/(Am) is vacuum permeability and current density

$$\mathbf{J} = \mathbf{J}_e + \mathbf{J}_m \quad (2)$$

(expressed in units of A/m²) is the sum of free-charge current density $\mathbf{J}_e$ and the equivalent current density, $\mathbf{J}_m = \nabla \times \mathbf{M}$, due to material of magnetization $\mathbf{M}$ (units of A/m). The sources of magnetic fields are therefore electric currents (for instance, in the Earth's core, the ionosphere or the magnetosphere) and/or magnetized material (for instance, in the Earth's crust).

Outside its sources (i.e., in regions with $\mathbf{J} = 0$), the magnetic field, $\mathbf{B} = -\nabla V$, is a Laplacian potential field and can be derived from a scalar magnetic potential V . In that case, the magnetic field has similar properties as the gravity field and the same methods for studying both fields might be used (Blakely, 1995). However, the similarity is only true under certain assumptions.

While near-surface magnetic field measurements are compatible with the potential assumption (since electric currents are negligible in the atmosphere and thus the data are obtained in a source-free region), the situation is different for spaceborne observations: Satellites probing the Earth's magnetic field typically fly at altitudes between 300 and 1,000 km in the ionospheric F -region, and especially at polar latitudes, the occurrence of local electric currents violates the assumption of a current-free data-sampling region. Satellite magnetic field data therefore cannot be described by a Laplacian potential field alone. Strategies for handling satellite data are discussed for instance in Hulot et al. (2007), Olsen et al. (2010), and Sabaka et al. (2010).

Applications of magnetic remote sensing

There are three main aspects of probing the Earth's interior using magnetic field observations.

Firstly, the dynamics of fluid flow at the top of the core can be determined from observations of the core field and its time change. This is a nonunique problem which requires assumptions on the flow dynamics, for instance to be steady or purely horizontal (i.e., toroidal). In addition to the fundamental nonuniqueness, the fact that only the slowly varying large-scale part of the core field is observable (periods shorter than a few months are heavily attenuated when propagating through the electrically conducting mantle and thus not observable at the surface, and the small-scale part of the core field is masked by the crustal field) puts additional limitations on the fraction of core flow that can be determined. Nevertheless, the magnetic field has proven to provide an extremely useful (and the only) tool to probe core flow. Recent overviews

on this subject are given by Holme (2007) and Whaler (2007).

Secondly, information on the structure, composition, and temperature of the crust can be obtained from investigations of the crustal magnetic field. Its small-scale part, mapped by regional near-surface surveys, provides information on shallow geological structures useful to petroleum and mineral exploration, while the long-wavelength part (up to 2,800 km – longer scales are masked by the core field) is a valuable tool to investigate the lower crust and upper mantle. See Langel and Hinze (1998), Purucker and Whaler (2007), and Ravat (2007) for more details on crustal field studies.

Finally, observations of the time-changing part of the magnetic field (of periods between seconds and several days) form the basis for electromagnetic induction studies, providing information on the electrical conductivity of the crust and mantle. The obtained conductivity distribution gives insight into geodynamic processes. This is complementary to seismic remote sensing since conductivity reflects the connectivity of constituents such as fluids, partial melt, and volatiles (all of which may have influence on rheology, mantle convection, and tectonic activity), while seismology ascertains bulk mechanical properties. A description of the techniques for probing the conductivity of the Earth's interior, and examples of obtained results, is given for instance in Constable (2007).

Conclusions

Magnetic field observations provide one of the few tools for remote sensing the Earth's interior, especially regarding the dynamics of core fluid flow and the structure, composition, and temperature of the crust. In addition, studying time variations of the magnetic field allows for determination of the electrical conductivity of the crust and mantle.

Bibliography

- Blakely, R. J., 1995. *Potential Theory in Gravity and Magnetic Applications*. Cambridge: Cambridge Press.
- Cain, J. C., 2007. POGO (OGO-2, -4 and -6 spacecraft). In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.
- Constable, S., 2007. Geomagnetic induction studies. In Kono, M. (ed.), *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 5, pp. 237–276.
- Finlay, C. C., Maus, S., Beggan, C. D., Bondar, T. N., Chambodut, A., Chernova, T. A., Chulliat, A., Golovkov, V. P., Hamilton, B., Hamoudi, M., Holme, R., Hulot, G., Kuang, W., Langlais, B., Lesur, V., Lowes, F. J., Luehr, H., Macmillan, S., Manda, M., McLean, S., Manoj, C., Menvielle, M., Michaelis, I., Olsen, N., Rauberg, J., Rother, M., Sabaka, T. J., Tangborn, A., Toffner-Clausen, L., Thébault, E., Thomson, A. W. P., Wardinski, I., Wei, Z., and Zvereva, T. I., 2010. International geomagnetic reference field: the eleventh generation. *Geophysical Journal International*, **183**, 1216–1230, doi:10.1111/j.1365-246X.2010.04804.x.

- Friis-Christensen, E., Lühr, H., and Hulot, G., 2006. Swarm: a constellation to study the earth's magnetic field. *Earth Planets Space*, **58**, 351–358.
- Friis-Christensen, E., Lühr, H., Hulot, G., Haegmans, R., and Purucker, M., 2009. Geomagnetic research from space. *EOS, Transactions, American Geophysical Union*, **900**(25), 213–215.
- Holme, R., 2007. Large-scale flow in the core. In Olson, P. (ed.), *Treatise on Geophysics*. Amsterdam: Elsevier Science, Vol. 8, pp. 107–130.
- Hulot, G., Sabaka, T. J., and Olsen, N., 2007. The present field. In Kono, M. (ed.), *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 5, pp. 33–75.
- Korhonen, J., Fairhead, J., Hamoudi, M., Hemant, K., Lesur, V., Manda, M., Maus, S., Purucker, M., Ravat, D., Sazonova, T., et al., 2007. *Magnetic Anomaly Map of the World*. Helsinki: Map published by Commission for Geological Map of the World, supported by UNESCO, GTK.
- Langel, R. A., and Hinze, W. J., 1998. *The Magnetic Field of the Earth's Lithosphere: The Satellite Perspective*. Cambridge: Cambridge University Press.
- Maus, S., 2007. CHAMP magnetic mission. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.
- Maus, S., Barckhausen, U., Berkenbosch, H., Bournas, N., Brozena, J., Childers, V., Dostaler, F., Fairhead, J., Finn, C., von Frese, R., et al., 2009. EMAG2: a 2-arc min resolution earth magnetic anomaly grid compiled from satellite, airborne, and marine magnetic measurements. *Geochemistry, Geophysics, Geosystems*, **10**(8), Q08005.
- Olsen, N., 2007. Ørsted. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.
- Olsen, N., Hulot, G., and Sabaka, T. J., 2010. Sources of the geomagnetic field and the modern data that enable their investigation, Chapter 5. In Freedman, W., Nashed, Z., and Sonar, T. (eds.), *Handbook of Geomathematics*. Heidelberg: Springer, pp. 106–124, doi:10.1007/978-3-642-01546-5_5.
- Purucker, M., and Whaler, K., 2007. Crustal magnetism. In Kono, M. (ed.), *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 5, pp. 195–235.
- Purucker, M. E., 2007. Magsat. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.
- Ravat, D., 2007. Crustal magnetic field. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.
- Roberts, P. H., 2007. Theory of the geodynamo. In *Treatise on Geophysics*. Amsterdam: Elsevier, Vol. 8, pp. 67–106.
- Sabaka, T. J., Hulot, G., and Olsen, N., 2010. Mathematical properties relevant to geomagnetic field modelling, Chapter 17. In Freedman, W., Nashed, Z., and Sonar, T. (eds.), *Handbook of Geomathematics*. Heidelberg: Springer, pp. 504–538, doi:10.1007/978-3-642-01546-5_17.
- Schmucker, U., 1985. Sources of the geomagnetic field. In *Landolt-Börnstein, New-Series, 5/2b*. Berlin/Heidelberg: Springer, pp. 31–73.
- Whaler, K. A., 2007. Core motions. In Gubbins, D., and Herrero-Bervera, E. (eds.), *Encyclopedia of Geomagnetism and Paleomagnetism*. Heidelberg: Springer.

Cross-references

Electromagnetic Theory and Wave Propagation
Fields and Radiation
Geodesy
Observational Systems, Satellite

MEDIA, ELECTROMAGNETIC CHARACTERISTICS

Yang Du

Zhejiang University, Hangzhou, People's Republic of China

Definition

Media, electromagnetic characteristics. Macroscopic permittivity and permeability properties of media.

Introduction

When electromagnetic wave propagates in some medium other than vacuum, since the characteristic wavelength is several orders larger than the atoms of which the medium is composed, the detailed behavior of the fields over atomic distance becomes irrelevant. What do matter are the quantities averaged over the atomic scale, including the macroscopic fields and macroscopic sources. Such treatment implies that the inhomogeneous medium is essentially replaced by a homogeneous one with macroscopic permittivity and permeability (Jackson, 1998).

Macroscopic properties

In macroscopic media, the electric displacement field **D** and magnetic field **H** are related to the electric field **E** and magnetic induction **B** through the macroscopically averaged electric dipole, magnetic dipole, electric quadrupole, and higher-order multipoles. In most materials, only the electric and magnetic polarizations are significant.

In an isotropic medium, the constitutive relations are **D** = ϵ **E**, and **B** = μ **H**, where ϵ and μ are the permittivity and permeability, respectively. A diamagnetic medium has μ larger than μ_0 , since diamagnetic substance, whose atoms or molecules have no angular momentum, creates induced magnetic moments that tend to oppose the applied magnetic field. A paramagnetic medium has μ smaller than μ_0 , since paramagnetic substance has a net angular momentum which is aligned parallel to the applied magnetic field.

For anisotropic media, the constitutive relations are **D** = $\bar{\epsilon}$ **E** and **B** = $\bar{\mu}$ **H**, where $\bar{\epsilon}$ and $\bar{\mu}$ are the permittivity and permeability tensor, respectively. Crystals are generally characterized by symmetric permittivity tensors, which can be transformed into a diagonal matrix as

$$\bar{\epsilon} = \begin{bmatrix} \epsilon_x & 0 & 0 \\ 0 & \epsilon_y & 0 \\ 0 & 0 & \epsilon_z \end{bmatrix}$$

When $\epsilon_x = \epsilon_y = \epsilon_z$, the crystals are isotropic, such as the cubic crystals. Uniaxial crystals have two of the three parameters equal. Examples are tetragonal, hexagonal, and rhombohedral crystals (Kong, 2005).

For bianisotropic media, the constitutive relations are **D** = $\bar{\epsilon}$ **E** + $\bar{\zeta}$ **H** and **B** = $\bar{\zeta}$ **E** + $\bar{\mu}$ **H** (Kong, 2005).

Dispersion

All media are dispersive in that the phase velocity of a wave depends on the frequency. That is, waves with different frequencies travel in a dispersive medium at different speeds. The frequency dependence of ϵ and μ leads to new effects when an arbitrary wave train containing a range of frequencies travels.

What enter the picture of dispersion are the molecular constitution of matter and the dynamics of molecules. A simple model leads to a dispersion formula which in most cases serves as an adequate representation of the dielectric constant as a function of frequency. The success of this model is truly amazing considering the several assumptions being made, including treating the relative permeability equal to unity, neglecting magnetic force effects, and confining the oscillation to be sufficiently small. The forces exerting upon an electron include an electric force due to the electric field, a restoring force, and a damping force. The resultant formula for the dielectric constant is

$$\epsilon(\omega) = \epsilon_0 + \frac{Ne^2}{m} \sum_k \frac{v_k}{\omega_k^2 - \omega^2 - i\omega\gamma_k},$$

where $-e$ and m are the charge and mass of an electron, respectively. Here it is assumed that there are N molecules per unit volume and v_k electrons per molecule with binding frequency ω_k and phenomenological damping constant γ_k (Jackson, 1998).

When real part of $\epsilon(\omega)$ increases with ω , it is called normal dispersion. Anomalous dispersion refers to the reverse. Since in general the resonant frequencies ω_k are large compared to the damping factors γ_k , in the neighborhood of ω_k , the behavior of $\epsilon(\omega)$ is expected to be rather violent, with disruption of the normal dispersion and appearance of appreciable imaginary part of $\epsilon(\omega)$ or resonant absorption.

At frequencies far beyond the highest resonant frequency, the dielectric constant is simply expressed in terms of the plasma frequency ω_p of the medium as $\epsilon(\omega) \approx \epsilon_0 - \frac{\omega_p^2}{\omega^2} \epsilon_0$.

The Kramers-Kronig relations or dispersion relations are simple integral formula relating the real part to imaginary part of the complex permittivity $\epsilon(\omega)$ or a dispersive process to an absorption process. The validity of these relations is very general due to the fact that very few assumptions are made, among which are the monochromatic components of the displacement $\mathbf{D}(\mathbf{r}, \omega)$ and the electric field $\mathbf{E}(\mathbf{r}, \omega)$ at position $\mathbf{r}$ related by $\mathbf{D}(\mathbf{r}, \omega) = \epsilon(\omega)\mathbf{E}(\mathbf{r}, \omega)$, as well as a causality requirement on the kernel relating $\mathbf{D}(\mathbf{r}, t)$ and $\mathbf{E}(\mathbf{r}, t)$. The wide generality of the dispersion relations makes them very useful in all areas of physics (Toll, 1956). For instance, in the scattering of nuclear particles, these relations connect real and imaginary parts of the diagonal elements of the scattering matrix (Jost et al., 1950).

Inhomogeneous composite

A macroscopically inhomogeneous medium is a typical representation of many materials. For instance, a snow pack is a composite of ice particles and water; a porous rock is a composite of the rock matrix and salt water if the latter is present. In each example, the medium possesses spatially varying quantities such as permittivity and conductivity.

Electromagnetic characterization of inhomogeneous composite materials is important in a wide variety of applications, such as in astronomy and atmospheric physics and selective absorbers of solar and infrared radiation.

In the effective medium approximation of Bruggeman and Landauer, a composite of two components is considered. To obtain the effective conductivity of this composite, each particle is imagined to be immersed in a homogeneous effective medium of conductivity σ_e instead of in its actual inhomogeneous medium. The self-consistency condition is invoked such that the average electric field within a particle shall equal the electric field far from the particle. In essence, in the effective-medium approximation, only electron-dipole scattering contributes. The characteristic particle dimension in the composite is required to be small compared to the characteristic wavelength. If a two-component composite has one component dominant in volume fraction, then both approaches work well. However, it is found that EMA produces significant error in far-infrared absorption by a composite composed of dielectric and small metal particles with concentrations below the percolation threshold (Stroud, 1998).

A self-consistency condition is to choose an effective dielectric constant such that for particles embedded in the medium, on the average, their forward scattering amplitude should vanish (Stroud and Pan, 1978). Generalization to continuous size distribution of particles as well as arbitrary number of components is provided in Chylek and Srivastava (1983), where both the contribution of electric and magnetic dipole terms are considered. Special attention is paid to a composite material where small metallic particles are among the components. The form of the size distribution of metallic particles is found to determine the critical volume fraction at which there is a drastic absorption increase.

Metamaterial

Metamaterial is a new class of artificially constructed electromagnetic material that can exhibit electromagnetic properties difficult or impossible to find in conventional materials (Pendry, 2004; Smith et al., 2004). For example, the material property of negative index of refraction has been realized at both GHz and optical frequencies using metamaterial concepts.

Recent advances in construction of metamaterial with independent and arbitrary varying capability of

permittivity and permeability values pave the way for totally new electromagnetic phenomena. For instance, the electric displacement $\mathbf{D}$, the magnetic field intensity $\mathbf{B}$, and the Poynting vector $\mathbf{S}$, by controlling the material electromagnetic properties, can be directed at will, made focused, or avoid objects. In particular is the design of cloaking of objects from electromagnetic fields (Pendry et al., 2006).

Summary

In macroscopic media, the electric displacement field and magnetic induction can be related to the electric intensity and magnetic intensity in ways quite different from that in vacuum. With increasing complexity, the medium can be characterized as isotropic, anisotropic, and bianisotropic. All media are dispersive in that the phase velocity of a wave depends on the frequency. The real and imaginary parts of the frequency-dependent complex permittivity are related by the Kramers-Kronig relations, which are very useful in all areas of physics. In characterizing the electromagnetic properties of an inhomogeneous medium, the effective-medium approximation is useful within its region of validity. Beyond that, the method base on the self-consistency condition can be employed. In terms of artificial materials, metamaterial is a new class of artificially constructed electromagnetic material that can exhibit electromagnetic properties difficult or impossible to find in conventional materials, which leads to totally new electromagnetic phenomena.

Bibliography

- Chylek, P., and Srivastava, V., 1983. Dielectric constant of a composite inhomogeneous medium. *Physical Review B*, **27**, 5098.
- Jackson, J. D., 1998. *Classical Electrodynamics*, 3rd edn. New York: Wiley.
- Jost, R., Luttinger, J. M., and Slotnick, M., 1950. Distribution of recoil nucleus in pair production by photons. *Physical Review*, **80**, 189.
- Kong, J. A., 2005. *Electromagnetic Wave Theory*. Cambridge: EMW.
- Pendry, J. B., 2004. A chiral route to negative refraction. *Science*, **306**, 1353.
- Pendry, J. B., Schurig, D., and Smith, D. R., 2006. Controlling electromagnetic fields. *Science*, **312**, 1780.
- Smith, D. R., Pendry, J. B., and Wiltshire, M. C. K., 2004. Metamaterials and negative refractive index. *Science*, **305**, 788.
- Stroud, D., 1998. The effective medium approximation: some recent developments. *Superlattices and Microstructures*, **23**, 567.
- Stroud, D., and Pan, F. P., 1978. Self-consistent approach to electromagnetic wave propagation in composite media: application to model granular metals. *Physical Review B*, **17**, 1602.
- Toll, J. S., 1956. Causality and the dispersion relation: logical foundations. *Physical Review*, **104**, 1760.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)

MICROWAVE DIELECTRIC PROPERTIES OF MATERIALS

Martti Hallikainen
Aalto University, Espoo, Finland

Synonyms

Dielectric constant; Permittivity

Definitions

Permittivity characterizes the electrical properties of materials: The real part gives the contrast with respect to vacuum and the imaginary part gives the electromagnetic loss of the material. Loss tangent is defined as the imaginary part divided by the real part of permittivity. Total electromagnetic loss in a medium consists of absorption loss (electromagnetic power transformed into other forms of energy, such as heat) and scattering loss (energy is caused to travel in directions other than that of incident radiation). Penetration depth provides an approximate value to the maximum depth of the medium that contributes to the backscattering coefficient and brightness temperature. Optical depth between two points in a medium provides a measure of how transparent the medium is to electromagnetic power passing through it. Both the real and imaginary parts of the permittivity of water are very large compared to other natural media; hence, the dielectric properties of water tend to dominate those of materials containing water.

Basic electromagnetic quantities for describing materials

Several basic electromagnetic quantities are defined in order to characterize the microwave interaction with matter. Two electrical quantities, electric flux density and electric field intensity, constitute a fundamental pair of electromagnetic fields, and they are related via

$$\vec{D} = \epsilon \epsilon_0 \vec{E} \quad (1)$$

where $\vec{D}$ is the electric flux density (displacement) vector, $\vec{E}$ is the electric field vector, ϵ is the relative (compared to vacuum) permittivity, and $\epsilon_0 = 8.854 \times 10^{-12}$ As/V/m is the vacuum permittivity. The permittivity of free space (air) is equal to that of vacuum with a high degree of accuracy.

Relative permittivity and loss tangent

The relative permittivity is a complex number and is denoted by

$$\epsilon = \epsilon' - j\epsilon'' \quad (2)$$

where $j = \sqrt{-1}$. It characterizes the electrical properties of materials: The real part (ϵ') gives the contrast with respect to free space ($\epsilon'_{air} = 1$), and the imaginary part (ϵ'') gives the electromagnetic loss of the material. The real

and imaginary parts of ε are often referred to as the dielectric constant and the dielectric loss factor, respectively. The loss tangent, defined as

$$\tan(\delta) = \frac{\varepsilon''}{\varepsilon'} \quad (3)$$

may also be used to characterize the loss of materials.

Propagation, absorption, and phase constant

For an electromagnetic plane wave traveling in the z -direction, the intensity of the electric field at point z can be expressed as

$$E(z) = E_0 e^{-\gamma z} \quad (4)$$

where E_0 is the field intensity at $z = 0$. The complex propagation constant of the medium is denoted by γ and is expressed as

$$\gamma = \alpha + j\beta \quad (5)$$

where α is the absorption constant and β is the phase constant. The absorption constant describes transformation of wave energy into other forms of energy, such as heat. The phase constant is equal to the wave number $k = 2\pi/\lambda$, where λ is wavelength, in a lossless medium. α and β are related to the complex permittivity by

$$\alpha = k_0 |\text{Im}\{\sqrt{\varepsilon}\}| \quad (6)$$

$$\beta = k_0 \text{Re}\{\sqrt{\varepsilon}\} \quad (7)$$

where k_0 is the wave number in free space.

Power absorption coefficient

The power absorption coefficient κ_a is defined as

$$\kappa_a = 2\alpha \quad (8)$$

and is often expressed in dB/m through the relation

$$\kappa_a(\text{dB/m}) = 8.686 \alpha (\text{Np/m}) \quad (9)$$

Extinction and scattering coefficient

The total electromagnetic loss in a medium consists of absorption loss (electromagnetic power transformed into other forms of energy, such as heat) and scattering loss (energy is caused to travel in directions other than that of the incident radiation). Scattering loss is caused by particles of different ε embedded in a host medium. The extinction coefficient (total loss) is thus

$$\kappa_e = \kappa_a + \kappa_s \quad (10)$$

where κ_s denotes the scattering loss.

Penetration depth

Part of a wave incident upon the surface of a medium from the air in the direction z is transmitted across the boundary

into the medium. The penetration depth δ_p is defined as the depth z at which (Ulaby et al., 1986)

$$\frac{P(z = \delta_p)}{P(z = 0+)} = \frac{1}{e} \quad (11)$$

where $P(\delta_p)$ is the transmitted power at depth δ_p and $P(0+)$ is the transmitted power just beneath the surface. If scattering in the medium is ignored, $\kappa_e \equiv \kappa_a = 2\alpha$ and if, additionally, α is constant as a function of z ,

$$\delta_p = \frac{1}{\kappa_a} = \frac{\sqrt{\varepsilon'}}{k_0 \varepsilon''}; \quad \varepsilon'' \ll \varepsilon' \quad (12)$$

The penetration depth provides an approximate value to the maximum depth of the medium that contributes to the backscattering coefficient and brightness temperature.

Optical depth

The optical depth between two points r_1 and r_2 in a medium provides a measure of how transparent the medium is to electromagnetic power passing through it. It is defined as

$$\tau(r_1, r_2) = \int_{r_1}^{r_2} \kappa_e dr \quad (13)$$

The optical depth is often used in atmospheric remote sensing.

Dielectric properties of natural media

The first book on the dielectric properties of materials was published by Debye, who presented his theory on polar liquids in 1929. Von Hippel published, in 1954, his book on theoretical and experimental aspects of dielectric media. The two-volume analysis of electric polarization was published by Böttcher (1973) and Böttcher et al. (1978). Followed by progress in microwave remote sensing in the 1970s, both theoretical and experimental studies of the dielectric properties of natural media increased rapidly. Electromagnetic mixing theories were analyzed by Sihvola (1999).

Electromagnetically, pure water is homogeneous matter, because its characteristics are not a function of the position within water. It is also isotropic, meaning that its characteristics are the same in all directions. However, most natural media are either nonhomogeneous or anisotropic; they may be mixtures of various substances, possibly with certain geometry. For example, soils consist of soil solids particles, air voids, and water. Seawater contains a variety of dissolved salts, making its dielectric behavior different from that of pure water. An example of anisotropic media is sea ice, which consists of ice, air, and mostly vertical brine inclusions.

In microwave remote sensing, interaction of electromagnetic waves with matter is modeled using the macroscopic dielectric properties of targets. Due to the

complex structure of many natural media, their dielectric characteristics cannot be derived theoretically with a sufficiently high degree of accuracy. Hence, the dielectric models for most media are semiempirical, relying on both theoretical studies and measurements made under laboratory conditions. In the following, the dielectric properties of natural media essential in microwave remote sensing are discussed. Various mixing models for computing the bulk dielectric properties of natural media are presented.

Most natural media contain water. Since the complex permittivity of water is substantially larger than that of other natural media at frequencies relevant to microwave radiometer and radar remote sensing, the dielectric properties of water tend to dominate those of other media including soils, vegetation, snow, and ice. The volumetric water content in soils, especially clay, may be up to 50 %, in vegetation parts up to 70 %, and in snow over 10 %. The permittivity of most natural dry materials is small ($\epsilon' < 5$ and $\epsilon'' < 0.1$). The dielectric properties of pure water, freshwater, and seawater are discussed first in this presentation, followed by those of other natural media important for microwave remote sensing.

Dielectric properties of pure water, freshwater, and seawater

Pure Water. For polar liquids like water, the complex permittivity can be derived by assuming that a polar molecule interacts with an electromagnetic field by rotating in a viscous medium. This results in the so-called Debye equation for water (Debye, 1929):

$$\epsilon_w = \epsilon_{w\infty} + \frac{\epsilon_{w0} - \epsilon_{w\infty}}{1 + j2\pi f\tau_w} \quad (14)$$

where ϵ_{w0} is relative permittivity of water in a static electric field, $\epsilon_{w\infty}$ is relative permittivity of water at an “infinitely high” frequency, τ_w is relaxation time of water, and f is frequency. ϵ'_w and ϵ''_w depend, in addition to frequency, on temperature. The real and imaginary part of permittivity can be solved from (Equation 14):

$$\epsilon'_w = \epsilon_{w\infty} + \frac{\epsilon_{w0} - \epsilon_{w\infty}}{1 + (2\pi f\tau_w)^2} \quad (15a)$$

$$\epsilon''_w = \frac{2\pi f\tau_w(\epsilon_{w0} - \epsilon_{w\infty})}{1 + (2\pi f\tau_w)^2} \quad (15b)$$

Numerical expressions for the terms of (Equation 14) as a function of temperature and salinity are available in Klein and Swift (1977); these equations have been regularly used for computing the complex permittivity of seawater. The frequency at which the maximum dielectric loss occurs is called the relaxation frequency; it is given by

$$f_0 = \frac{1}{2\pi\tau_w} \quad (16)$$

It has been recently concluded that the model based on the Debye relaxation process (Equation 14) is accurate to within less than 1 % up to 30 GHz at temperatures around 25 °C, but differences between the extrapolated value and independently measured permittivity values are about 10 % at 90 GHz (Ellison, 2006).

There are two approaches to increase accuracy at frequencies beyond 30 GHz and at low temperatures, the double Debye model of Liebe et al. (1991) and the extended double Debye model of Ellison (2006). The double Debye model is basically the same as that given in (Equation 14), but it includes two relaxation processes instead of one (Liebe et al., 1991):

$$\epsilon_w = \frac{\epsilon_{w01} - \epsilon_{w\infty1}}{1 + j\left(\frac{f}{f_{01}}\right)} + \frac{\epsilon_{w\infty1} - \epsilon_{w\infty2}}{1 + j\left(\frac{f}{f_{02}}\right)} + \epsilon_{w\infty2} \quad (17)$$

where ϵ_{w01} is the static permittivity, f_{01} and f_{02} are the first and second relaxation frequency, respectively, and $\epsilon_{w\infty1}$ and $\epsilon_{w\infty2}$ are the first and second high-frequency permittivity, respectively, and

$$\theta = 1 - \frac{300}{T(K)} \quad (18)$$

$$\epsilon_{w01} = 77.66 - 103.3\theta \quad (19)$$

$$\epsilon_{w\infty1} = 0.0671\epsilon_{w01} \quad (20)$$

$$f_{01} = 20.20 + 146.4\theta + 316\theta^2 \quad (21)$$

$$\epsilon_{w\infty2} = 3.52 + 7.52\theta \quad (22)$$

$$f_{02} = 39.8f_{01} \quad (23)$$

where $T(K)$ is the temperature in Kelvins.

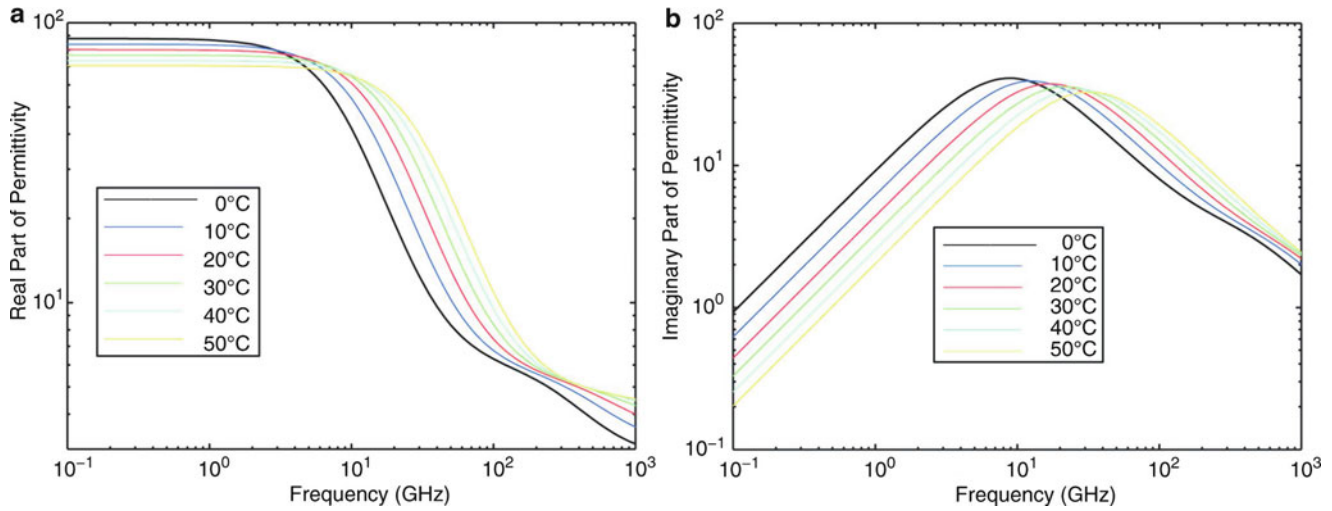
Based on the double Debye model in (Equations 17–23), Figure 1 shows the complex permittivity of pure water over the frequency range of 100 MHz–1,000 GHz for temperatures between 10 °C and 50 °C.

Freshwater. There are relatively few dissolved ions in freshwater, and hence, ϵ' for freshwater is practically the same as that for pure water. However, the loss term must be corrected by adding a conductivity term:

$$\epsilon''_{iw} = \epsilon''_w + \frac{\sigma}{2\pi f\epsilon_0} \quad (24)$$

where subscript i refers to freshwater (impure water) and σ is the conductivity. Due to the frequency term in the denominator of (Equation 24), the effect of conductivity increases with decreasing frequency.

Pure Water and Seawater. Seawater contains dissolved salts, including NaCl, MgCl₂, Na₂SO₄, CaCl₂, and KCl, and the average salinity in the oceans is 32.5 psu (Anderson, 1960). Dissolved salts in seawater increase the ionic conductivity. Consequently, the dielectric loss factor of seawater is substantially higher than that of pure water at low microwave frequencies.



Microwave Dielectric Properties of Materials, Figure 1 Complex dielectric permittivity of pure water according to Liebe et al. (1991).

For this reason, L-band is optimal for retrieval of ocean salinity from radiometer data.

Recently, the dielectric behavior of water has been approximated using an extended double Debye model up to 25 THz for pure water and up to 105 GHz for seawater (Ellison, 2006). His results are based on several data sets and indicate that there are three relaxation processes in the microwave region and two resonances in the far-infrared region. At 25 °C, the relaxation frequencies are 18.56 GHz, 167.83 GHz, and 1.944 THz; the two resonance frequencies are 4.03 and 14.48 THz. Freshwater and seawater contain dissolved ions and gases; hence, their dielectric behavior is different from that of pure water. Using the extended double Debye model, the complex permittivity is then expressed (Ellison, 2006) as

$$\epsilon_w = \epsilon_{w\infty} + \frac{\epsilon_{w0} - \epsilon_{w1}}{1 + j2\pi f \tau_{w1}} + \frac{\epsilon_{w1} - \epsilon_{w\infty}}{1 + j2\pi f \tau_{w2}} - j \frac{\sigma_w}{2\pi f \epsilon_0} \quad (25)$$

The parameters ϵ_{w0} , ϵ_{w1} , $\epsilon_{w\infty}$, τ_{w1} , τ_{w2} , and σ_w depend on temperature and salinity of water. Their highly complex numerical expressions are given in (Ellison, 2006, pp. 445–454):

$$\epsilon_s(T, S) = 87.85306 \exp(-0.00456992T - a_1S - a_2S^2 - a_3ST) \quad (26)$$

$$\epsilon_1(T, S) = a_4 \exp(-a_5T - a_6S - a_7ST) \quad (27)$$

$$\tau_1(T, S) = (a_8 + a_9S) \exp\left(\frac{a_{10}}{T + a_{11}}\right) \quad (28)$$

$$\tau_2(T, S) = (a_{12} + a_{13}S) \exp\left(\frac{a_{14}}{T + a_{15}}\right) \quad (29)$$

$$\epsilon_{\infty}(T, S) = a_{16} + a_{17}T + a_{18}S \quad (30)$$

$$\sigma(T, S) = \sigma(T, 35) \cdot P(S) \cdot Q(T, S) \quad (31)$$

where

$$\sigma(T, 35) = 2.903602 + 8.607 \cdot 10^{-2}T + 4.738817 \cdot 10^{-4}T \quad (32)$$

$$- 2.991 \cdot 10^{-6}T^3 + 4.3041 \cdot 10^{-9}T^4 \quad (33)$$

$$P(S) = S \frac{37.5109 + 5.45216S + 0.014409S^2}{1,004.75 + 182.283S + S^2} \quad (34)$$

and

$$Q(T, S) = 1 + \frac{\alpha_0(T - 15)}{T + \alpha_1} \quad (35)$$

where

$$\alpha_0 = \frac{6.9431 + 3.2841S - 0.099486S^2}{84.85 + 69.024S + S^2} \quad (36)$$

$$\alpha_1 = 49.843 - 0.2276S + 0.00198S^2 \quad (37)$$

Equations 25–37 with Table 1 are stated to represent the permittivity of pure water to within 1 % over the frequency range 0–20 GHz, to within 3 % over the frequency range 30–100 GHz, and to within 5 % over the frequency range 100–1,000 GHz (Ellison, 2006). The permittivity of seawater is represented to within 3 % over the frequency

Microwave Dielectric Properties of Materials, Table 1 Coefficients for Equations 26–30

$a_1 = 0.46606917$ E (–2)	$a_7 = 0.34414691$ E (–4)	$a_{13} = 0.38957681$ E (–6)
$a_2 = -0.26087876$ E (–4)	$a_8 = 0.17667420$ E (–3)	$a_{14} = 0.30742330$ E (3)
$a_3 = -0.63926782$ E (–5)	$a_9 = -0.20491560$ E (–6)	$a_{15} = 0.12634992$ E (3)
$a_4 = 0.63000075$ E (1)	$a_{10} = 0.58366888$ E (3)	$a_{16} = 0.37245044$ E (1)
$a_5 = 0.26242021$ E (–2)	$a_{11} = 0.12634992$ E (3)	$a_{17} = 0.92609781$ E (–2)
$a_6 = -0.42984155$ E (–2)	$a_{12} = 0.69227972$ E (–4)	$a_{18} = -0.26093754$ E (–1)

range 3–105 GHz within temperatures of 0–30 °C and salinities of 0–40 psu. Below 3 GHz, the results of Klein and Swift (1977) can be used.

Dielectric properties of snow

Dry Snow. Dry snow consists of ice crystals and air voids. Any water in the snow medium collects at points of contact between the ice grains. The metamorphism caused by melting and freezing changes the microstructure of snow. The ice grains become rounded during the melting process and some of the smaller grains disappear completely. Snow that has undergone several melt–freeze cycles tends to form multiple clusters. In general, the density of snow slowly increases with time due to metamorphism and melt–freeze cycles. The density of dry snow varies from 0.1 g/cm³ (newly fallen snow) to 0.5 g/cm³ (refrozen snow).

Since dry snow is a dielectric mixture of ice and air, its complex permittivity is governed by the dielectric properties of ice, snow density, and ice-particle shape. Since the real part of the permittivity of ice is $\varepsilon'_i = 3.17$ at frequencies between 10 MHz and 1,000 GHz (Mätzler and Wegmüller, 1987) and practically independent of temperature, the dielectric constant of dry snow, ε'_{ds} is only a function of density. For the most extensive data set used in dielectric measurements of snow, the result between 3 and 37 GHz is (Hallikainen et al., 1986)

$$\varepsilon'_{ds} = 1 + 1.9 \rho_{ds}; \quad \rho_{ds} \leq 0.5 \text{ g/cm}^3 \quad (38)$$

where ρ_{ds} is the density of dry snow in g/cm³. From (Equation 38), $\varepsilon'_{ds} = 1.57$ for a density of 0.3 g/cm³. A piecewise linear relationship between ε'_{ds} and ρ_{ds} has been reported based on measurements at 1 GHz and related modeling (Mätzler, 1996).

Experimental data for the loss factor of dry snow are limited to frequencies below 13 GHz (Cumming, 1952; Tiuri et al., 1984). The best fit to the experimental data of Tiuri et al. was provided by the following expression:

$$\varepsilon''_{ds} = 1.59 \cdot 10^6 (0.52 \rho_{ds} + 0.62 \rho_{ds}^2) \cdot \left(f^{-1} + 1.23 \cdot 10^{-14} \sqrt{f} \right) \exp(0.036T) \quad (39)$$

where T is the temperature in °C and f is the frequency in Hz. An alternative equation has been provided by Ulaby et al., (1986) using a two-phase dielectric mixing formula for spherical inclusions. Both approaches indicate that the dielectric loss factor of dry snow has a minimum value between 1 and 5 GHz and then increases with increasing frequency. It is of the order of 10^{-4} – 10^{-3} at all frequencies below 100 GHz; hence, dry snow is a loss–loss medium.

Due to its granular structure, dry snow is a strongly scattering medium at high microwave frequencies. Based on theoretical calculations using the Mie theory (assuming independent scattering), scattering dominates over absorption in dry snow (ice grain size 1 mm) at frequencies above 15 GHz (Ulaby et al., 1986). Scattering loss increases rapidly with increasing frequency; hence, the extinction coefficient at and above 35 GHz consists mainly of scattering loss. Based on an extensive data set of snow samples ranging from newly fallen to refrozen snow, with the ice grain diameter ranging from 0.2 to 1.6 mm and snow density from 0.17 to 0.39 g/cm³, the following empirical expressions have been developed to relate the extinction coefficient of dry snow, denoted κ_{eds} , to the observed snow particle diameter d_s (Hallikainen et al., 1987):

$$\kappa_{eds} = 1.5 + 7.4 d_s^{2.3} \text{ dB/m at 18 GHz} \quad (40)$$

$$\kappa_{eds} = 30 d_s^{2.1} \text{ dB/m at 35 GHz} \quad (41)$$

$$\kappa_{eds} = 180 d_s^{2.0} \text{ dB/m at 60 GHz} \quad (42)$$

$$\kappa_{eds} = 300 d_s^{1.9} \text{ dB/m at 90 GHz} \quad (43)$$

The particle diameter (observed by photography) is in millimeters. Equations 40–43 hold for particle sizes below 1.6 mm. Equations 40–42 can be combined into a single equation of the form

$$\kappa_{eds} = 0.0018 f^{2.8} d_s^{2.0} \text{ dB/m} \quad 18 \text{ GHz} < f < 60 \text{ GHz} \quad (44)$$

where f is in GHz and d_s is in mm. Due to the exponent value of 2.8 in (Equation 44), the extinction coefficient increases rapidly with increasing frequency; at 35 GHz it is about 20 dB/m for snow with a grain size of 1 mm, whereas at 90 GHz it is well over 100 dB/m. This is why radiometers operating at 35 GHz are able to measure snow water equivalents less than 100 mm and those operating at 90 GHz “see” only the snowpack surface. The strong fluctuation theory (Stogryn, 1986) provides results that agree reasonably well with the empirical extinction coefficient values in the 18–60 GHz range for all realistic grain sizes and also at 90 GHz for grain sizes smaller than 0.9 mm.

Wet Snow. Wet snow is a mixture of ice crystals, liquid water, and air. The geometry and porosity of wet snow depend on its liquid water content. It has been concluded that snow has two distinct regimes of liquid saturation (Colbeck, 1982). In the lower range (pendular regime), air is continuous throughout the pore space, and liquid water occurs in the form of isolated inclusions. In the higher range of liquid saturation (funicular regime), liquid water is continuous throughout the pore space, and air occurs as distinct bubbles trapped by narrow constrictions in the pores. There is a sharp transition between the two regimes.

Electromagnetically, wet snow is a three-component dielectric mixture consisting of ice particles, air, and liquid water. Both water and ice exhibit Debye-type relaxation spectra. The relaxation frequency (frequency at which the maximum dielectric loss occurs) of ice is in the kilohertz range, whereas that for water at 0 °C is 9 GHz. The complex permittivity of ice and water depends on frequency and temperature. Consequently, the permittivity of wet snow is a function of frequency, temperature, volumetric water content, snow density, ice-particle shape, and the shape of water inclusions.

Since the permittivity of water is substantially higher than that of ice and air, the dielectric behavior of wet snow is governed by the volume fraction of water. Basically, ϵ'_{ws} and ϵ''_{ws} are compressed versions of those for water at 0 °C. The dielectric behavior of wet snow was measured by Hallikainen et al. (1986) at nine frequencies between 3 and 18 GHz and, additionally, at 37 GHz. They tested four dielectric models: Debye-like model, modified Debye-like model, two-phase Polder–Van Santen model (Polder and Van Santen 1946), and three-phase Polder–Van Santen model. The two-phase Polder–Van Santen mixing formula is applied by assuming wet snow to consist of dry snow as the host material with water inclusions embedded in it. The shape of the water inclusions was included in the modeling study. Tiuri et al. (1984) used a resonator in the 0.5–1 GHz range and also the 4 GHz data set of Hallikainen et al. (1982) and developed a numerical model for the permittivity of wet snow. Mätzler et al. (1984) made resonator measurements at 1 GHz and applied their results for spectral modeling of wet snow. See Hallikainen et al. (1986) for a review of dielectric wet snow studies.

The samples of Hallikainen et al. (1986) had densities ranging from 0.09 to 0.42 g/cm³ and liquid water contents ranging from 0 % to 12.3 % by volume. The snow particle size varied between 0.5 and 1.5 mm and the sample temperatures ranged from –5 °C to 0 °C. The two-phase Polder–Van Santen model for wet snow is

$$\epsilon_{ws} = \epsilon_{ds} + \frac{m_v \epsilon_{ws}}{3} (\epsilon_w - \epsilon_{ds}) \cdot \sum_{j=1}^3 [\epsilon_{ws} + (\epsilon_w - \epsilon_{ws}) A_{wj}]^{-1} \quad (45)$$

where m_v is the volumetric liquid water content and water droplets are randomly distributed and randomly oriented

ellipsoids with depolarization factors A_{w1} , A_{w2} , and A_{w3} . The general Debye-like model is expressed as

$$\epsilon'_{ws} = A + \frac{Bm_v^x}{1 + (f/f_0)^2} \quad (46a)$$

$$\epsilon''_{ws} = \frac{C(f/f_0)m_v^x}{1 + (f/f_0)^2} \quad (46b)$$

where f_0 is the relaxation frequency (9.07 GHz) and A , B , C , and x are constants determined by fitting the model to the measured data in the 3–37 GHz range. The obtained values are given in Hallikainen et al. (1986). In the simplified Debye-like model, A , B , and C depend only on dry snow density and volumetric water content, whereas in the modified Debye-like model, they depend, additionally, on frequency. The Debye-like equations do not take into account the geometry of the wet snow medium (shape of water inclusions, funicular and pendular regimes).

Liquid water inclusions in snow are approximately needle shaped in the pendular regime (low values of water content), but they become approximately disk shaped in the funicular regime (high values of water content) (Colbeck, 1982). The best fit of the developed two-phase Polder–Van Santen mixing model to experimental data was obtained by assuming that their shapes depend on snow water content and, additionally, that the water inclusions are nonsymmetrical in shape (Hallikainen et al., 1986). Transition from the pendular regime to the funicular regime was observed to take place around snow volumetric water content $m_v = 3$ %.

Both the modified Debye-like model and the two-phase Polder–Van Santen model (with variable, nonsymmetrical shape factors) provide good accuracy (Hallikainen et al., 1986). The simplified Debye-like model works well only at frequencies below 15 GHz, but the modified Debye-like model works well also at higher frequencies. The two-phase Polder–Van Santen model with wetness-dependent shape factors provides a slightly better overall fit to the experimental observations of wet snow. The modified Debye-like model is easier to use and is likely to provide adequate accuracy for most applications. The three-phase Polder–Van Santen model provided accuracies comparable to those of the two-phase Polder–Van Santen model.

The real part of the dielectric constant of wet snow varies from the values for dry snow up to 3.0 at 3 GHz and to 1.9 at 37 GHz for a liquid water content of 12 % and, generally, decreases with increasing frequency. Its dielectric loss factor increases with frequency until it reaches its maximum value of 0.9 at 9 GHz for a wetness of 12 %. At frequencies above 9 GHz, it decreases monotonically. Even a volumetric water content of 2 % results in absorption of tens of decibels per meter, depending on frequency. The penetration depth for wet snow is generally less than 20 cm at frequencies above 5 GHz, suggesting that microwave radar and radiometer measurements can provide information only on the top of the snowpack.

Dielectric properties of ice

Pure and Freshwater Ice. Freshwater ice is free from salt, but usually it includes air bubbles, impurities, and cracks. Pure ice is an idealization of freshwater ice (frozen distilled water); it is not discussed here. The dielectric properties of freshwater ice at microwave frequencies are determined by two physical processes, namely, the high-frequency tail of a relaxation spectrum (relaxation frequency in the kilohertz range) and the low-frequency tail of far-infrared absorption bands (Evans, 1965; Mätzler and Wegmüller, 1987). These processes result in a practically constant value for the dielectric constant ϵ'_i at microwave frequencies and in a minimum of the dielectric loss factor ϵ''_i centered around 1–5 GHz.

Based on experimental data, the relative dielectric constant of pure and impure ice may be assigned the constant value of

$$\epsilon'_i = 3.17 \quad (47)$$

for frequencies between 10 MHz and 100 GHz (Mätzler and Wegmüller, 1987; Cumming, 1952).

For the imaginary part of the permittivity, results of various investigators show substantial scatter. The scatter is caused, in addition to measurement inaccuracies (partly due to small values of ϵ''_i), by varying amounts of impurities in the ice samples. Reviews of results are available, for example, in Ulaby et al. (1986). A numerical equation for ϵ''_i was developed by Mätzler and Wegmüller (1987):

$$\epsilon''_i = \frac{A}{f} + Bf^C \quad (48)$$

where f is the frequency in GHz. The values of constants A , B , and C were determined for pure and impure ice at temperatures -5°C and -15°C . The values for impure ice at -5°C are $A = 0.0026$, $B = 0.00023$, and $C = 0.87$. For pure ice, they are $A = 0.0006$, $B = 0.000065$, and $C = 1.07$. For -15°C , they are for impure ice $A = 0.0013$, $B = 0.00012$, and $C = 1.0$. For pure ice, they are $A = 0.00035$, $B = 0.000036$, and $C = 1.2$. Based on (Equation 41), ϵ''_i for impure ice is between 0.001 and 0.01 in the 1–100 GHz range and slightly smaller for pure ice. Thus, freshwater ice is a low-loss medium, and its power absorption coefficient at 1 GHz (temperature -5°C) is about 0.1 dB/m and penetration depth is 100 m. At 10 GHz, the corresponding numbers are about 1 dB/m and 10 m. This means that contributions to the brightness temperature and the backscattering coefficient are not limited to the topmost ice layers.

Sea Ice. Sea ice consists of freshwater ice, liquid brine, and air. In addition to the dielectric properties of these constituents, several additional parameters influence its complex permittivity, including the volume fraction of each constituent and geometry (shape, size, and orientation) of brine pockets with respect to the propagation direction of the electromagnetic wave. Most of these parameters depend on ice temperature. The salinity of liquid brine in

sea ice is governed by its temperature, and it is over 200 ‰ by weight at -20°C and decreases to about 35 ‰ at -2°C (Assur, 1960). The relative volume of liquid brine is directly proportional to the salinity; for sea ice with a salinity of 1 ‰, it is about 3 ‰ at -20°C , 24 ‰ at -2°C , and 103 ‰ at -0.5°C (Frankenstein and Garner, 1967). Due to the high volumetric content and salinity, the complex permittivity of brine is substantially higher than that of seawater (Stogryn and Desargant, 1985). Hence, brine characteristics dominate the permittivity of sea ice.

Experimental investigations of the microwave dielectric behavior of sea ice have been carried out at frequencies below 40 GHz. Reviews of these investigations are given in Vant (1976) and Hallikainen and Winebrenner (1992). The total number of permittivity measurements for arctic ice is small. Additionally, most of the investigations are limited to first-year ice. Frequency-wise, extensive sets of experimental Arctic sea ice data were acquired by Vant (1976), Vant et al. (1978) (0.1–7.5 GHz; 13 samples), and Sackinger and Byrd (1972) (26–40 GHz; sample number not reported). Sample-wise, the most extensive studies are those of Bogorodskii and Khokhlov (1975) (10 GHz; 130 samples) and Hallikainen (1983) (0.6 and 0.9 GHz; 373 samples). References Vant (1976) and Vant et al. (1978) report the same results, but the former includes detailed results for each ice sample. Hoekstra and Cappillino (1971) (0.1–24 GHz) measured ice samples of various salinities, made from seawater flash frozen in a sample holder; their results have very likely been influenced by sample preparation technique.

In most studies, temperatures below -5°C were used, although rapid changes in the permittivity occur at temperatures close to 0°C due to increasing relative volume of brine. Hallikainen (1983) measured low-salinity ($S < 2$ ‰) ice samples at temperatures up to -0.1°C . In order to avoid brine drainage, each ice sample was measured immediately after removing it from the ice field and using only the original temperature. The relative brine volume of low-salinity sea ice near 0°C may be higher than that of high-salinity sea ice at lower temperatures. The relative brine volume of sea ice of salinity 1 ‰ is 103 ‰ at -0.5°C , whereas the corresponding number for sea ice of salinity 4 ‰ is 40 ‰ at -5°C .

A limited amount of data is available on the dielectric properties of multiyear sea ice (Vant, 1976, two samples; Vant et al., 1978, two samples). Due to lower salinity, the complex permittivity of multiyear ice is lower than that of first-year ice.

Only Vant et al. (1978) (NaCl ice; orientation from vertical 0° , 30° , 45° , 60° , and 90°), Sackinger and Byrd (1972) (Arctic first-year ice; 0° and 90°), and Hallikainen (1983) (low-salinity first-year ice; 0° , 30° , and 90°) have investigated the effect of sample orientation with respect to the propagation direction of the electromagnetic wave. Sackinger and Byrd (1972) observed that the dielectric loss factor at -7°C may be up to 300 % higher for 90° (horizontal orientation) than for 0° (vertical); it decreases

with decreasing temperature and is about 150 % higher at -21.5°C . Hence, the dielectric loss of sea ice increases with increasing incidence angle. Based on their data, Vant et al. (1978) developed a model to account for sample orientation. The permittivity also depends on ice type; in general, the dielectric loss is higher for frazil than for columnar sea ice with the electromagnetic wave propagating in the vertical direction.

Comparisons between various data sets have shown that there is substantial scatter for values of both the real part and imaginary part of the permittivity for mostly used frequencies like 1 GHz, 5 GHz, and 10 GHz (Hallikainen and Winebrenner, 1992). Efforts to model the dielectric behavior of sea ice are scarce. Vant et al. (1978) used two approaches based on the relative brine volume and a semiempirical approach to account for the orientation. Stogryn (1987) employed strong fluctuation theory to compute the tensor dielectric constant for sea ice. Hallikainen and Winebrenner (1992) generated numerical equations, based on available data sets, to relate the dielectric properties of sea ice to the relative brine volume V_b at three frequencies – 1 GHz, 4 GHz, and 10 GHz. Their expressions are given below in a slightly modified form, due to revisiting the data sets.

$$\varepsilon'_{si} = 3.12 + 0.009 V_b; \quad 1 \text{ GHz} \quad (49a)$$

$$\varepsilon''_{si} = 0.04 + 0.005 V_b; \quad 1 \text{ GHz} \quad (49b)$$

$$\varepsilon'_{si} = 3.05 + 0.007 V_b; \quad 4 \text{ GHz} \quad (50a)$$

$$\varepsilon''_{si} = 0.02 + 0.0033 V_b; \quad 4 \text{ GHz} \quad (50b)$$

$$\varepsilon'_{si} = 3.04 + 0.004 V_b; \quad 10 \text{ GHz} \quad (51a)$$

$$\varepsilon''_{si} = 0.01 + 0.007 V_b; \quad 10 \text{ GHz} \quad (51b)$$

where the relative brine volume $V_b \leq 70\%$ and is given by (Frankenstein and Garner, 1967):

$$V_b = S \left(-\frac{52.56}{T} - 2.28 \right) \quad -0.5^{\circ}\text{C} \geq T \geq -2.06^{\circ}\text{C} \quad (52a)$$

$$V_b = S \left(-\frac{45.917}{T} + 0.930 \right) \quad -2.06^{\circ}\text{C} > T \geq -8.2^{\circ}\text{C} \quad (52b)$$

$$V_b = S \left(-\frac{43.795}{T} + 1.189 \right) \quad -8.2^{\circ}\text{C} > T \geq -22.9^{\circ}\text{C} \quad (52c)$$

Equations 49a, b–51a, b do not explicitly take into account the effect of ice density upon the real part of the permittivity or direction of wave propagation in the ice medium. The correlation coefficient for Equations 49a, b–51a, b is, in general, between 0.7 and 0.8.

In general, the absorption coefficient κ_a increases with increasing frequency and temperature. The loss of multiyear ice is considerably lower than that of first-year ice, mainly due to much lower salinity. The values for first-year low-salinity sea ice show that the temperature dependence of the absorption coefficient is high at temperatures near 0°C . At 10 GHz, the differences between the results for columnar, frazil, and multiyear ice are substantial.

In the 1–10 GHz range, the penetration depth is between 100 and 5 cm for first-year ice and 500 and 30 cm for multiyear ice. Sensors operating at X-band provide information on sea ice mainly from the topmost 5–80 cm, depending on ice type, salinity, and temperature, whereas the corresponding numbers for L-band sensors are 40–500 cm.

Dielectric properties of soils

Soils consist of bulk soil, air, and water. A soil's textural composition is usually given in terms of weight percentages of sand, silt, and clay. Sand includes particles with diameters in the range between 0.05 and 2.0 mm, silt includes particles with diameters in the 0.002–0.05 mm range, and clay includes particles with diameters smaller than 0.002 mm. The water contained in the soil is usually divided into bound water and free water, although the transition between these is not sharp. Molecules of bound water are contained in the first few molecular layers surrounding the soil particles; these are tightly held by the soil particles due to the matric and osmotic forces (Baver et al., 1977). The amount of bound water is proportional to the total surface area of the soil particles, which depends on the soil particle size distribution and mineralogy: The smaller the particles, the more there can be bound water. Hence, clay has more bound water than sand. Water molecules located further away from the soil particle surface can move within the soil medium and are referred to as free water.

From the electromagnetic point of view, soil is a heterogeneous medium consisting of bulk soil, air, bound water, and free water. Its dielectric properties as a function of frequency depend on the soil bulk density (compaction), soil composition (particle size distribution and mineralogy), the volume fractions of bound and free water, the salinity of the soil solution, and temperature (Dobson et al., 1985). Wang and Schmugge (1980) showed that various dielectric mixing formulas describing soil as a two-component system (soil with free water) fail to describe the complex permittivity of wet soil realistically at 1.4 and 5 GHz. They were the first to explicitly include both free water and bound water in their empirical mixing formula by examining two moisture regions: (a) water contents less than the maximum bound water fraction and (b) water contents higher than the bound water fraction. They included two free parameters in the model and, by optimizing their values, were able to explain the behavior of the complex permittivity, especially its real part. Wang (1980) modeled the soil–water system with

a Debye-like relaxation over a finite band of relaxation frequencies and, by adjusting two free parameters, adequately predicted the behavior of data over the 0.3–1.4 GHz range. The two free parameters were the width of the activation energy of the soil solution and the mean relaxation frequency of the soil–water mixture at a given frequency and for a given soil and water content.

Hallikainen et al. (1985) performed dielectric measurements of five soil types at 12 frequencies between 1.4 and 18 GHz. Soil moisture values from 0 % to the highest moisture contents that can be supported by that soil type without drainage taking place were used; additionally, data for frozen soils were acquired as well. They developed for each frequency polynomial expressions, separately for real and imaginary parts of the permittivity, dependent on volumetric moisture content and the percentage of sand and clay contained in the soil. Dobson et al. (1985) developed two dielectric models for wet soil based on the data set of Hallikainen et al. (1985), a theoretical model and a semiempirical model. The theoretical model accounts explicitly for the presence of bound water adjacent to hydrophilic soil particle surfaces and employs a four-component dielectric mixing model that describes the soil–water system consisting of dry soil solids as a host medium with randomly distributed and randomly oriented disc-shaped inclusions of bound water, bulk water, and air. The bulk water component characteristics in the model depend on frequency, temperature, and salinity. Based on comparisons with data, the theoretical model was determined to be an appropriate formulation, and it yields values that describe well the observed effects of frequency and soil type. However, its accuracy is limited by the uncertainty concerning the dielectric properties of bound water.

The driving force in the development of the semiempirical model of Dobson et al. (1985) was to develop a user-friendly, frequency-dependent model that is based on readily measured soil characteristics including volumetric moisture and weight fractions of sand and clay. The final expressions for the real and imaginary parts of the permittivity are

$$\epsilon'_{ws} = \left[1 + \frac{\rho_b}{\rho_s} (\epsilon_s^\alpha - 1) + m_v^{\beta'} \epsilon_{fw}^{\alpha'} - m_v \right]^{\frac{1}{\alpha}} \quad (53a)$$

$$\epsilon''_{ws} = \left[m_v^{\beta''} \epsilon_{fw}^{\alpha''} \right]^{\frac{1}{\alpha}} \quad (53b)$$

where α is a constant shape factor, ρ_b is soil bulk density, $\rho_s = 2.66 \text{ g/cm}^3$ is soil specific density, ϵ_s is the soil solid permittivity, m_v is the volumetric moisture, β' and β'' are the soil-texture-dependent coefficients for computing the real and imaginary part, respectively, and ϵ_{fw} is the permittivity of free water. The optimum value for the shape factor was determined to be $\alpha = 0.65$. The relative permittivity of soil solids is

$$\epsilon_s = (1.01 + 0.44\rho_s)^2 - 0.062 \quad (54)$$

and the soil-texture-dependent coefficients are

$$\beta_{\epsilon'} = 0.01(127.48 - 0.519S - 0.152C) \quad (55a)$$

$$\beta_{\epsilon''} = 0.01(1.33797 - 0.603S - 0.166C) \quad (55b)$$

where S and C are the percentage of sand and clay, respectively. The dielectric properties of free water are computed using an expression similar to that for saline water:

$$\epsilon_w = \epsilon_{w\infty} + \frac{\epsilon_{w0} - \epsilon_{w\infty}}{1 + j2\pi f \tau_w} - j \frac{\sigma_{eff}}{2\pi f \epsilon_0} \cdot \frac{\rho_s - \rho_b}{\rho_s m_v} \quad (56)$$

where the effective conductivity is

$$\sigma_{eff} = -1.645 + 1.939\rho_b - 0.02013S + 0.01594C \quad (57a)$$

The above expressions were determined using the whole soil dielectric data set between 1.4 and 18 GHz. For the frequency range of 0.3–1.3 GHz, another expression for the effective conductivity was developed (Peplinski et al., 1995):

$$\sigma_{eff} = 0.0467 + 0.2204\rho_b - 0.4111S + 0.66144C \quad (57b)$$

For a reader interested in the dielectric properties of soils at a certain frequency between 1.4 GHz and 18 GHz, the polynomial expressions in Hallikainen et al. (1985) are useful. Mironov et al. (2004) developed a generalized refractive index mixing dielectric model for moist soils incorporating the dielectric characteristics of bound water.

The above studies show that the complex permittivity of wet soil is a compressed version of that for slightly saline water. In the 1.4–18 GHz range, the real part of the permittivity increases with increasing volumetric moisture and decreasing frequency, whereas the imaginary part increases with increasing moisture and increasing frequency. At any given moisture content and at all frequencies, the real part is roughly proportional to sand content and inversely proportional to clay content; however, this effect decreases with increasing frequency. The effect of soil texture on the imaginary part varies with frequency. At 1.4 GHz, the real part increases with increasing clay content for moisture levels above $0.2 \text{ cm}^3/\text{cm}^3$, whereas at 4–6 GHz, it is nearly independent of soil texture at all moisture levels and at frequencies of 8 GHz and above it decreases with increasing clay content. This behavior is obviously due to the ionic conductivity (strongest at low frequencies) and the relation of volume fraction of bound water to soil specific surface. At moisture levels close to saturation, ϵ may reach values up to 23-j3 for sand ($m_v = 0.35$) and 33-j9 for clay ($m_v = 0.50$) at 1.4 GHz. At 18 GHz, the corresponding values are 13-j8 and 18-j12, respectively. The penetration depth for wet soil decreases drastically with increasing moisture content and frequency;

consequently, retrieval of soil moisture from microwave radiometer and radar data works best at low frequencies. Even at 1.4 GHz, information only on surface moisture is obtained.

The permittivity of soils decreases drastically at temperatures below 0 °C (Hallikainen et al., 1985). Between −11 °C and −24 °C, both the real and imaginary parts of permittivity still depend on temperature, demonstrating that not all water is frozen. Recently, Mironov et al., (2010) conducted dielectric measurements on organic rich permafrost soil from 1 to 16 GHz and from −30 °C to +25 °C. They also extended the previously developed dielectric soil model of Mironov et al. (2004) to cover frozen soil.

Dielectric properties of vegetation

The characteristics of plants and trees vary substantially, including the size, density, moisture content, and geometry. Hence, the physical characteristics of vegetation material are not discussed in this presentation. For a detailed discussion on timber characteristics, the reader is referred to Dinwoodie (2000). Dielectric properties of wood are discussed in detail in Torgovnikov (1993); they have been primarily examined for industrial purposes, especially for microwave drying of timber at 2.45 GHz.

The dielectric properties of plants including trunks, stalks, and leaves for remote sensing have been measured and modeled by El-Rayes and Ulaby (1987), and Ulaby and El-Rayes (1987) in the 0.2–20 GHz range, and monitored in situ at L-band by McDonald et al. (1999). Mätzler (1994) measured and modeled various leaves at frequencies 1–94 GHz. Ulaby et al. (1987) measured the propagation constant for vegetation canopies with vertical stalks between 1.6 and 10.2 GHz. Dielectric spectroscopy of Scots pine at frequencies up to 1 GHz has been done by Tomppo et al. (2009).

Ulaby and El-Rayes (1987) modeled the complex permittivity of vegetation as a simple additive mixture of plant material, free water, and bound water due to their observation that bound water has a substantially lower relaxation frequency than free water. At 22 °C, their Debye–Cole dual-dispersion model for the permittivity of vegetation is

$$\begin{aligned} \varepsilon_v = \varepsilon_r + v_{fw} \left[4.9 + \frac{75.0}{1 + j \cdot 0.556f} - j \frac{18\sigma}{f} \right] \\ + v_{bw} \left[2.9 + \frac{55.0}{1 + (j \cdot 0.556f)^{0.5}} \right] \end{aligned} \quad (58)$$

where ε_r is the permittivity of bulk vegetation material, f is the frequency in GHz, σ is the ionic conductivity of the free-water solution in S/m, and v_{fw} and v_{bw} are the relative fractions of free and bound water, respectively. The volumetric moisture content consists of free water and bound water:

$$M_u = v_{fw} + v_{bw} = \frac{M_g \rho}{1 - M_g(1 - \rho)} \quad (59)$$

where M_g denotes the gravimetric moisture content and ρ is the bulk density. If M_g and ρ are known, the following expressions can be used for obtaining higher accuracy:

$$\varepsilon_r = 1.7 + 3.2M_v + 6.5M_v^2 \quad (60)$$

$$v_{fw} = M_v(0.82M_v + 0.166) \quad (61)$$

$$v_{bw} = \frac{31.4M_v^2}{1 + 59.5M_v^2} \quad (62)$$

The Debye–Cole dual-dispersion model has been determined to provide an estimate of ε'_v for leaves, stalks, branches, and trunks with an accuracy of 5 %. If the salinity S is known in order to provide an estimate for the conductivity (S/m) via

$$\sigma = 0.16S - 0.0013S^2 \quad (63)$$

similar accuracies apply to ε''_v .

Summary

Dielectric properties of natural media at microwave frequencies have been reviewed, and the effects of various physical parameters to the behavior of their permittivity have been discussed. Since most natural media are heterogeneous, theoretical derivation of their dielectric properties is not straightforward. Measurements provide useful information for model development and evaluation, but only within the range of the covered physical characteristics and employed frequencies. The permittivity of pure and saline water is the key to modeling the behavior of most natural media; gaps still occur for its values at near-freezing temperatures, especially at high microwave frequencies. More experimental data are needed to confirm the behavior of wet snow at frequencies above 18 GHz. Measurements of the permittivity of sea ice for various salinities and ice types, especially for high brine volumes and at temperatures above −5 °C, would substantially benefit modeling of the behavior of sea ice. Permittivity measurements of various types of vegetation are needed to test presently available semiempirical models and further develop them.

Bibliography

- Anderson, D. L., 1960. The physical constants of sea ice. *Research*, **13**, 310–318.
- Assur, A., 1960. *Composition of Sea Ice and Its Tensile Strength*. Wilmette: U.S. Army Snow, Ice and Permafrost Research Establishment.
- Baver, L. D., Gardner, W. H., and Gardner, W. R., 1977. *Soil Physics*. New York: Wiley.
- Bogorodskii, V. V., and Khokhlov, G. P., 1975. Electrical properties of ice in the ice edge zone of the Bering Sea at 10 GHz frequency. In *Proceedings of the Final Symposium on the Results of the Joint Soviet-American Expedition*. Leningrad: Gidrometeoizdat, pp. 219–233 (in Russian).
- Böttcher, C. J. F., 1973. *Theory of Electrical Polarisation*, 2nd edn. Amsterdam: Elsevier, Vol. 1.

- Böttcher, C. J. F., Van Belle, O. C., Bordewijk, P., and Rip, A., 1978. *Theory of Electrical Polarisation*, 2nd edn. Amsterdam: Elsevier, Vol. 2.
- Colbeck, S. C., 1982. The geometry and permittivity of snow at high frequencies. *Journal of Applied Physics*, **53**, 4495–4500.
- Cumming, W., 1952. The dielectric properties of ice and snow at 3.2 cm. *Journal of Applied Physics*, **23**, 768–773.
- Debye, P., 1929. *Polar Molecules*. New York: Dover.
- Dinwoodie, J. M., 2000. *Timber, Its Nature and Behaviour*, 2nd edn. New York: Taylor & Francis.
- Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., and El-Rayes, M., 1985. Microwave dielectric behavior of wet soil – part II: dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing*, **23**, 35–46.
- Ellison, W., 2006. Microwave dielectric properties of water. In Mätzler, C., Rosenkranz, P. W., Battaglia, A., and Wigneron, J. P. (eds.), *Thermal Microwave Radiation: Applications for Remote Sensing*. Stevenage: Institute of Engineering and Technology. IET Electromagnetic Waves Series, Vol. 52, pp. 431–455.
- El-Rayes, M. A., and Ulaby, F. T., 1987. Microwave dielectric spectrum of vegetation – part I: experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 541–549.
- Evans, S., 1965. Dielectric properties of ice and snow – a review. *Journal of Glaciology*, **5**, 773–792.
- Frankenstein, G., and Garner, R., 1967. Equation for determining the brine volume of sea ice from -0.5°C to -22.9°C . *Journal of Glaciology*, **6**, 943–944.
- Hallikainen, M. T., 1983. A new low-salinity sea-ice model for UHF radiometry. *International Journal of Remote Sensing*, **4**, 655–681.
- Hallikainen, M. T., and Winebrenner, D., 1992. The physical basis for sea ice remote sensing. In Carsey, F. (ed.), *Microwave Remote Sensing of Sea Ice*. American Geophysical Union, Geophysical Monograph 68, Chap. 3, Washington, DC, pp. 29–46.
- Hallikainen, M. T., Ulaby, F. T., and Abdel-Razik, M., 1982. Measurements of the dielectric properties of snow in the 4–18 GHz frequency range. In *Proceedings of the 12th European Microwave Conference*. Helsinki, pp. 151–156.
- Hallikainen, M. T., Ulaby, F. T., Dobson, M. C., and El-Rayes, M., 1985. Microwave dielectric behavior of wet soil – part I: empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, **23**, 25–34.
- Hallikainen, M. T., Ulaby, F. T., and Abdelrazik, M., 1986. Dielectric properties of snow in the 3 to 37 GHz range. *IEEE Transactions on Antennas and Propagation*, **34**, 1329–1340.
- Hallikainen, M. T., Ulaby, F. T., and Van Deventer, T. E., 1987. Extinction behavior of dry snow in the 18 to 90 GHz range. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 737–745.
- Hoekstra, P., and Cappillino, P., 1971. Dielectric properties of sea and sodium chloride ice at UHF and microwave frequencies. *Journal of Geophysical Research*, **76**, 4922–4931.
- Klein, L. A., and Swift, C. T., 1977. An improved model for the dielectric constant of seawater at microwave frequencies. *IEEE Transactions on Antennas and Propagation*, **25**, 104–111.
- Liebe, H. J., Hufford, G. A., and Manabe, T., 1991. A model for the complex permittivity of water at frequencies below 1 THz. *International Journal of Infrared and Millimeter Waves*, **12**, 659–675.
- Mätzler, C., 1994. Microwave (1–100 GHz) dielectric model of leaves. *IEEE Transactions on Geoscience and Remote Sensing*, **32**, 947–949.
- Mätzler, C., 1996. Microwave permittivity of snow. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 573–581.
- Mätzler, C., and Wegmüller, U., 1987. Dielectric properties of fresh-water ice at microwave frequencies. *Journal of Physics D: Applied Physics*, **20**, 1623–1630. Errata, 1988. 21:1660.
- Mätzler, C., Aebischer, H., and Schanda, E., 1984. Microwave dielectric properties of wet snow. *IEEE Journal of Oceanic Engineering*, **9**, 366–371.
- McDonald, K. C., Zimmermann, R., Way, J. B., and Chun, W., 1999. Automated instrumentation for continuous monitoring of the dielectric properties of woody vegetation: system design, implementation, and selected in situ measurements. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 1880–1894.
- Mironov, V. L., De Roo, R. D., and Savin, I. V., 2010. Temperature-dependable microwave dielectric model for an Arctic soil. *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 2544–2556.
- Mironov, V. L., Dobson, M. C., Kaupp, V. H., Komarov, S. A., and Kleshchenko, V. N., 2004. Generalized refractive mixing dielectric model for moist soils. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 773–785.
- Peplinski, N. R., Ulaby, F. T., and Dobson, M. C., 1995. Dielectric properties of soils in the 0.3–1.3 GHz range. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 803–807. Correction: 33: 1340.
- Polder, D., and Van Santen, J. H., 1946. The effective permeability of mixtures of solids. *Physica*, **12**, 257–271.
- Sackinger, W. M., and Byrd, R. C., 1972. *Reflection of Mm waves from snow and sea ice*, IAEA Report 7203, Institute of Arctic Environmental Engineering, University of Alaska, Fairbanks.
- Sihvola, A., 1999. *Electromagnetic Mixing Formulas and Applications*. London: Institution of Electrical Engineers. IEE Electromagnetic Wave Series, Vol. 47.
- Stogryn, A., 1986. A study of the microwave brightness temperature of snow from the point of view of strong fluctuation theory. *IEEE Transactions on Geoscience and Remote Sensing*, **24**, 220–231.
- Stogryn, A., 1987. An analysis of the tensor dielectric constant of sea ice at microwave frequencies. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 147–158.
- Stogryn, A., and Desargeant, G. J., 1985. The dielectric properties of brine in sea ice at microwave frequencies. *IEEE Transactions on Antennas and Propagation*, **33**, 523–532.
- Tiuri, M., Sihvola, A., Nyfors, E., and Hallikainen, M. T., 1984. The complex dielectric constant of snow at microwave frequencies. *IEEE Journal of Oceanic Engineering*, **9**, 377–382.
- Tomppo, L., Tiitta, M., Laakso, T., Harju, A., Venäläinen, M., and Lappalainen, R., 2009. Dielectric spectroscopy of scots pine. *Wood Science and Technology*, **43**, 653–667.
- Torgovnikov, G. I., 1993. *Dielectric Properties of Wood and Wood-Based Materials*. Berlin: Springer.
- Ulaby, F. T., and El-Rayes, M. A., 1987. Microwave dielectric spectrum of vegetation – part II: dual-dispersion model. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 550–557.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1986. *Microwave Remote Sensing – Active and Passive, III: From Theory to Applications*. Dedham: Artech House.
- Ulaby, F. T., Tavakoli, A., and Senior, T. B. A., 1987. Microwave propagation constant for a vegetation canopy with vertical stalks. *IEEE Transactions on Geoscience and Remote Sensing*, **25**, 714–725.
- Vant, M. R., 1976. *A Combined Empirical and Theoretical Study of the Dielectric Properties of Sea Ice over the Frequency Range 100 MHz to 40 GHz*. Technical Report, Carleton University, Ontario, Ottawa.
- Vant, M. R., Ramseier, R. O., and Makios, V., 1978. The complex-dielectric constant of sea ice at frequencies in the range 0.1–40 GHz. *Journal of Applied Physics*, **49**, 1264–1280.
- Von Hippel, A., 1954. *Dielectrics and Waves*. Boston: MIT Press.
- Wang, J. R., 1980. The dielectric properties of soil-water mixtures at microwave frequencies. *Radio Science*, **15**, 977–985.
- Wang, J. R., and Schmugge, T. J., 1980. An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Transactions on Geoscience and Remote Sensing*, **18**, 288–295.

MICROWAVE HORN ANTENNAS

Yahya Rahmat-Samii

Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA, USA

Definition

Horn Antenna. An antenna with metallic flare walls which provides a transition between waves propagating in a waveguide and electromagnetic waves directivity radiated into space.

Microwave horn antennas

Horns are one of the simplest and most widely used microwave antennas. Interests in horn antennas date back to the turn of the nineteenth century and then considerably revived during World War II. Microwave horn antennas are essentially a device to make a transition from waves propagating in a waveguide into electromagnetic signals transmitting in another open medium such as free space. They can have metallic walls, dielectric material, or the combination of both, and occur in various shapes and sizes to fulfill many practical applications, such as communication systems, remote sensing, radio frequency heating, reference sources for other antenna testing and evaluation. Other than being a stand-alone directive power transmitter/receiver, horns are also used as feeds for other antennas such as reflectors, lenses, and compound antennas. Their widespread application is due to their simple, solid geometry and excellent performance when beam directivity is required (Love, 1976; Balanis, 1988, 1996; Olver et al., 1994; Bird and Love, 2007).

Figure 1 summarizes the primary categories of horns, and Figure 2 shows some representative examples of well-known horn designs. Rectangular pyramidal horns

are probably the most commonly used class of horns. One of the primary applications of a pyramidal horn is the standard gain antenna since its gain may be calculated very accurately by knowing their dimensions. The beamwidths in the two principal planes can also be independently controlled by varying the rectangular-aperture dimensions. Some other forms of rectangular horns with nonlinear profiles are also used to increase aperture efficiency and lower side lobes compared with a linear flare. Circular geometries are also in widespread use. The axial symmetry of conical horns allows them to generate any polarization of the dominant mode which makes them well suited for circular polarization. The beamwidths, however, are usually unequal in the two principal planes. In order to overcome this problem, the dual-mode and corrugated (hybrid-mode) conical horns have been developed. Higher-order modes are excited so that the aperture field of the horn is modified in such a way as to produce radiation patterns with axial symmetry and very low cross-polarization.

Horn specifications

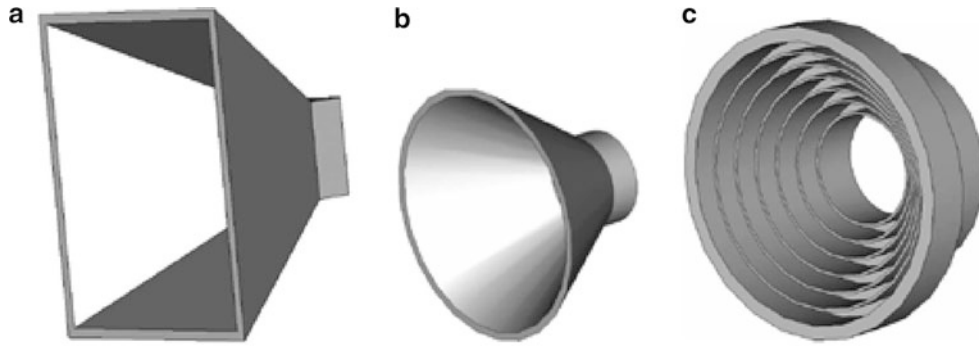
Ample types of horns can provide various radiation performances, and for every specific application, it is necessary for an antenna designer to consider some critical specifications which dictate the choosing of an appropriate type of horn antennas.

Bandwidth

The frequency band over which the system is to operate is usually specified as a range of frequencies (for single-band applications) where the antenna must satisfy a required return loss (or VSWR) and provide a radiation pattern adequate for the application. Consider an antenna required to operate from $f_{\min}$ to $f_{\max}$. The bandwidth, in general, can be calculated as

Horn Antennas		
Rectangular	Circular	Other Types
Smooth wall	Smooth-wall conical	Diagonal
Corrugated	Corrugated	Elliptical
Dielectric-lined	Scalar	Dielectric
Profiled	Dielectric-lined	
Multimode	Coaxial	
Ridged	Profiled	
	Longitudinal corrugation	
	Multimode	
	Ridged	

Microwave Horn Antennas, Figure 1 A summary of primary types of horn antennas.



Microwave Horn Antennas, Figure 2 Representative types of horn antennas: (a) pyramidal horn, (b) circular horn, and (c) corrugated horn.

$$\text{Bandwidth (\%)} = \frac{2(f_{\max} - f_{\min})}{(f_{\max} + f_{\min})} \times 100.$$

Based on this definition, a bandwidth below 20 % is generally considered narrowband whereas a bandwidth over 20 % is considered wideband. Often, there is a specific frequency determined in specifications as the design frequency at which the design provides optimum performances. For single-band applications, a practical choice is a design frequency of $f_c = \sqrt{f_{\min} \cdot f_{\max}}$ for narrowband operations and $f_c = 1.2 f_{\min}$ for wideband operations.

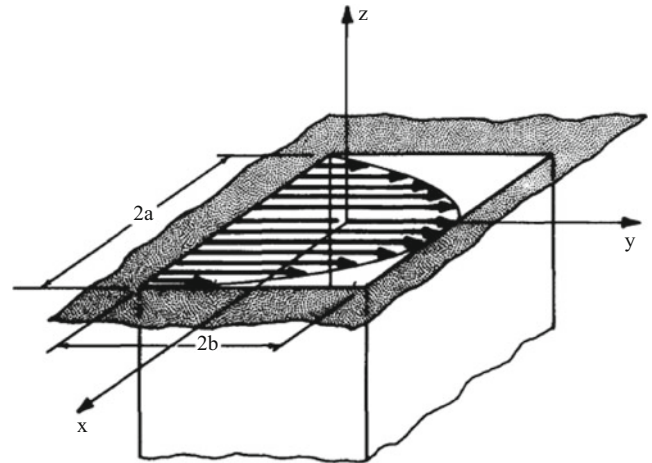
Gain and aperture efficiency

The antenna gain takes into account both the directional capability and efficiency of the antenna. Among many different types of gains, two definitions are usually used to describe horn antennas (Balanis, 1996). The most commonly accepted definition of gain only considers the accepted power and removes the power reflected from the input. In this definition, an ideal horn without losses is assumed and the gain equals directivity as the efficiency is 100 %. Another definition, called absolute gain, retains the reflected power from the input, as this is an intrinsic property of the antenna. For horn antennas, the results of these two definitions are usually very similar since the reflection coefficient is small. The difference is a factor of $1/(1 - |\Gamma_{\text{in}}|^2)$, where Γ_{in} is the reflection coefficient at the input.

Another useful measure for comparing the performances of horn antennas is the aperture efficiency. It is defined as the ratio of the antenna gain in the direction (θ, ϕ) and the gain of the same antenna aperture with ideal uniform illumination given by

$$\eta_{\text{aperture}} = \frac{\lambda^2}{4\pi A} G(\theta, \phi),$$

where A is the total area of the aperture, λ is the wavelength at the operating frequency, and $G(\theta, \phi)$ is the gain function. In many applications it is assumed that the aperture efficiency is in the boresight direction of the antenna.

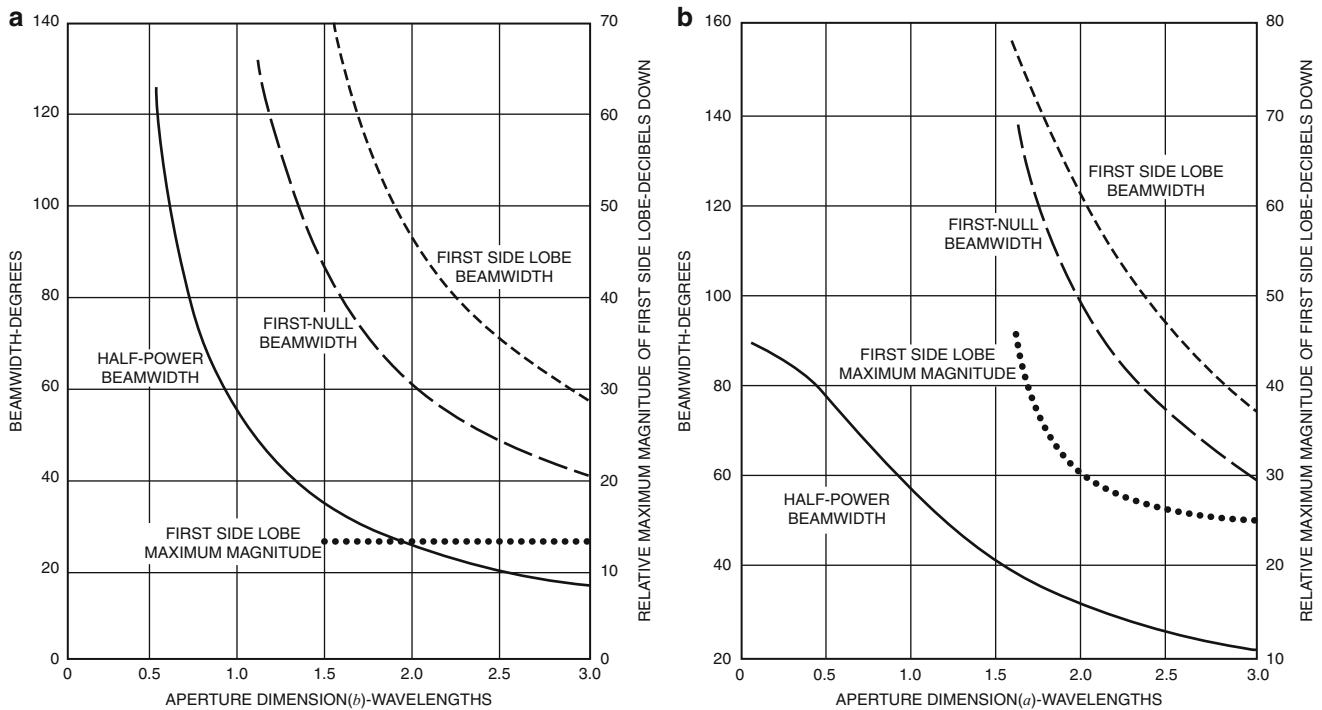


Microwave Horn Antennas, Figure 3 A TE_{10} -mode open-ended rectangular waveguide feed with dimensions $2a$ and $2b$.

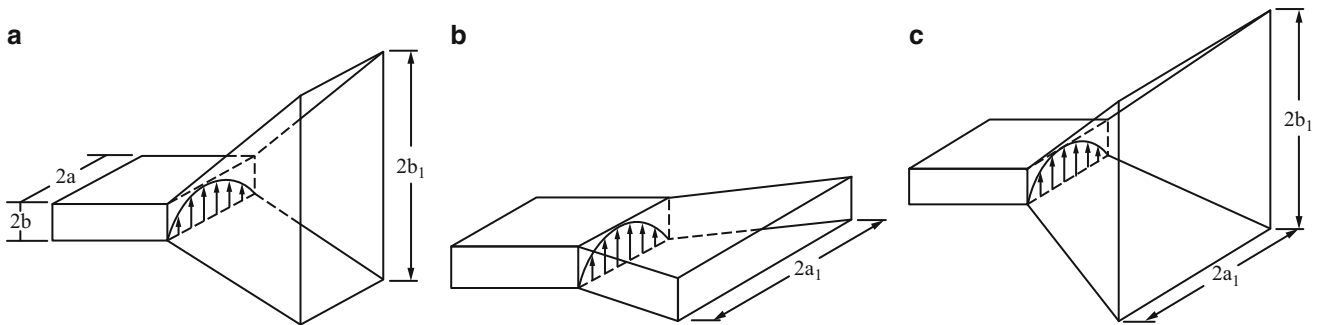
It has to be noted that what is usually defined is the gain of the whole antenna system, not just the gain of the aperture antenna in isolation. Therefore, all other losses such as mismatch and ohmic losses have to be taken into consideration. A well-designed aperture antenna fed by a poorly designed feed system results in poor performance.

Polarization

The radiated far-field of a horn antenna can be decomposed into two components: The one in the direction of the desired polarization is called co-polarization, and the orthogonal component is called cross-polarization. Polarization performance measurement can determine, in general, how much power is radiated in the undesired polarization which cannot be received by receiving antennas. The most common measurement is the peak cross-polarization level for linear polarization and the axial ratio for circular polarization. The measurement of the cross-polarization can be achieved in different ways. For linearly polarized horns, the co-polar pattern is measured by aligning the test antenna with the



Microwave Horn Antennas, Figure 4 Different beamwidths and first side lobe levels for TE₁₀-mode open-ended rectangular waveguide on ground plane: (a) E-plane and (b) H-plane.



Microwave Horn Antennas, Figure 5 Typical rectangular horn antennas: (a) E-plane sectoral horn, (b) H-plane sectoral horn, and (c) pyramidal horn.

polarization of the distant source antenna, and the cross-polar pattern is obtained by rotating the source antenna by 90° and repeating the radiation pattern measurement.

Functioning environment

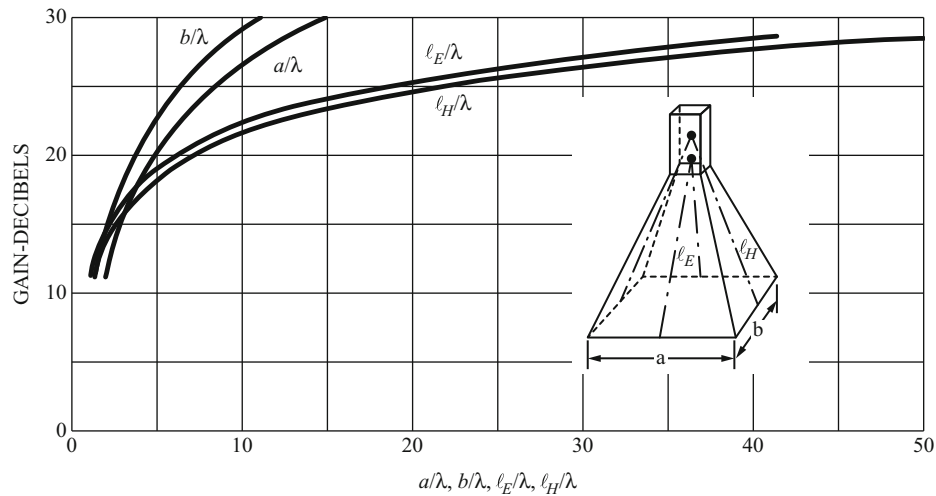
Consideration of the working environment in which the antenna is likely to operate is important in design process as well. In many cases this too will dictate the choice of materials and the mechanical aspects of the system. One consideration is the potential effects of surrounding objects on the horn's performance. A horn antenna may work perfectly well in isolation but give poor performance when surrounded by nearby scatterers. A typical effect is

caused by conductive surrounding objects which are fitted afterward in forms of blockage or radome effect. Other considerations are any special needs for horn antennas such as pressurization, high-power handling, low passive intermodulation level, robustness, and low-loss requirements.

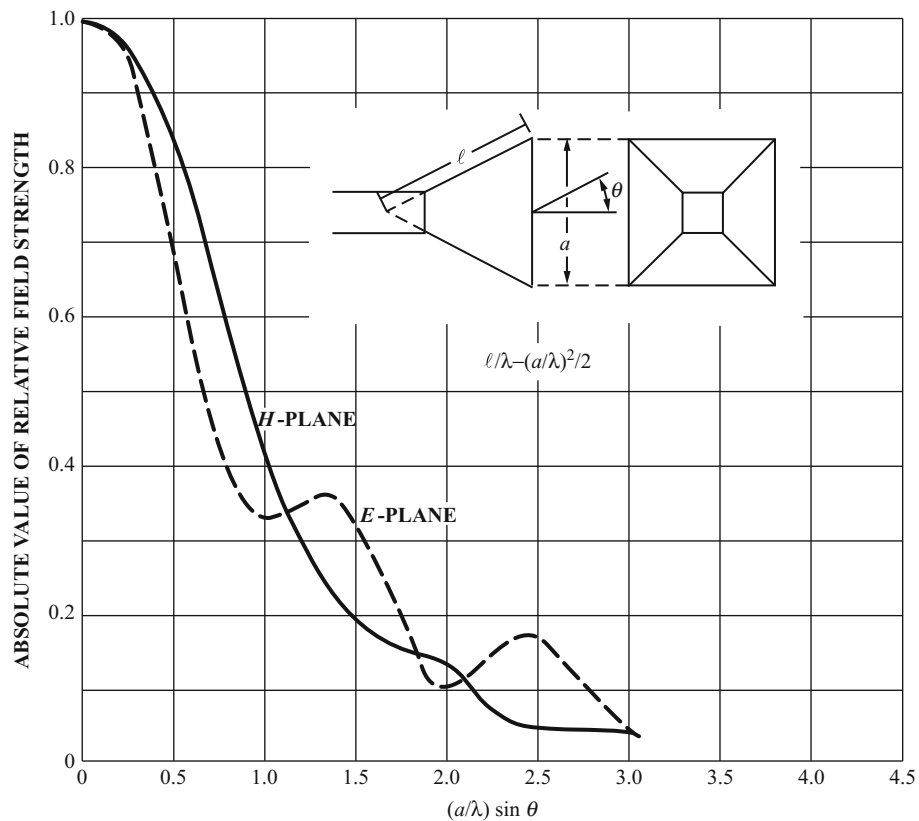
Conventional horn antennas

Rectangular horn

Rectangular horns are constructed by gradually flaring a rectangular waveguide to provide a smooth transition from the input to free space. If the flare angle is small enough so that the higher-order modes can be



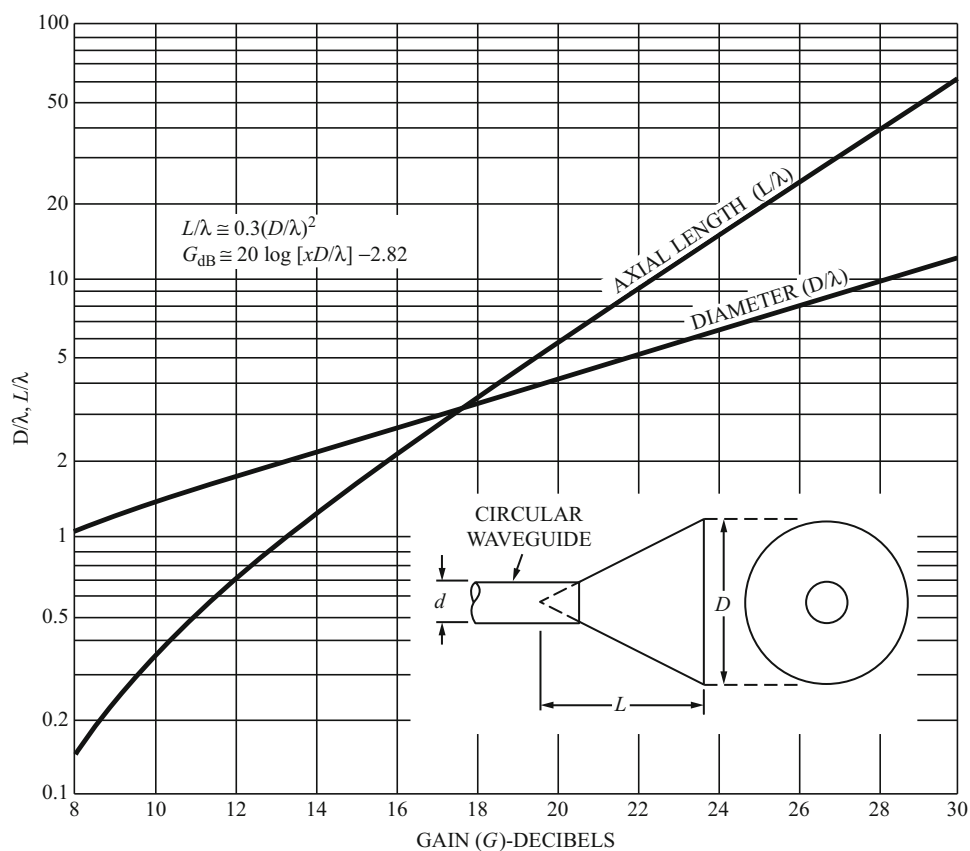
Microwave Horn Antennas, Figure 6 Gain characteristics of pyramidal horns of optimum-gain design.



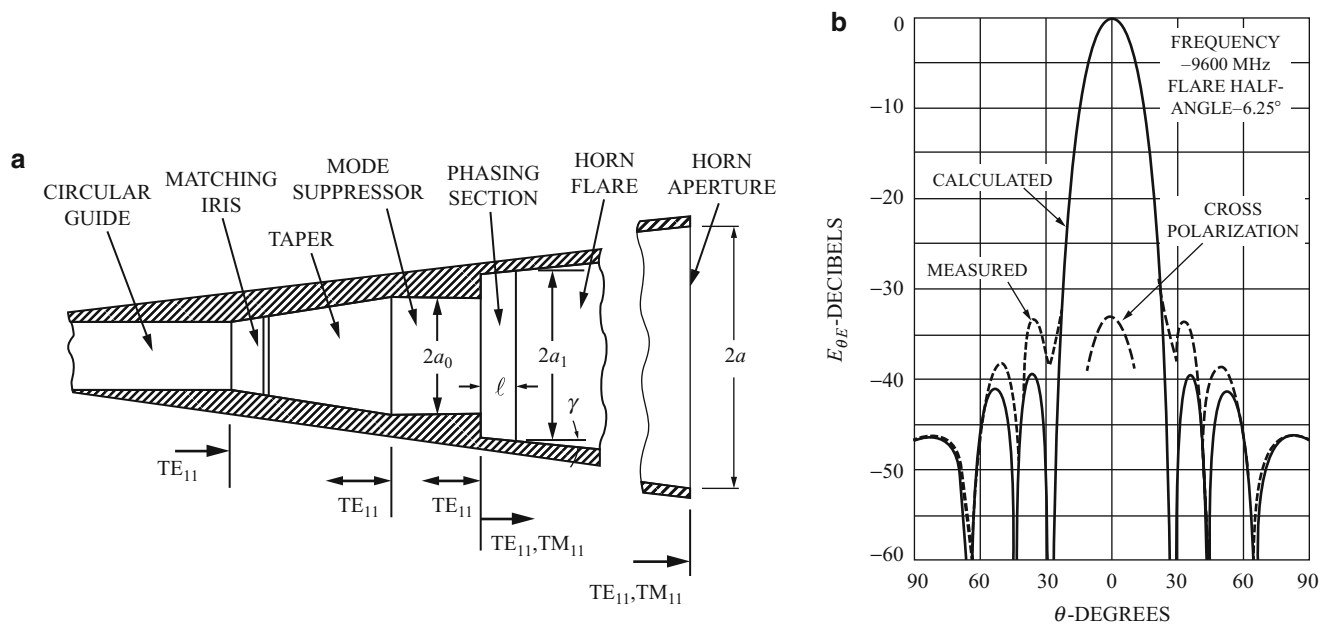
Microwave Horn Antennas, Figure 7 Universal patterns of pyramidal horns of optimum-gain design.

neglected, the transverse electric field of the dominant mode of the waveguide is a good first approximation to the aperture field at the horn mouth. For reference, some important characteristics of the open-ended rectangular waveguide (Figure 3) are shown in Figure 4.

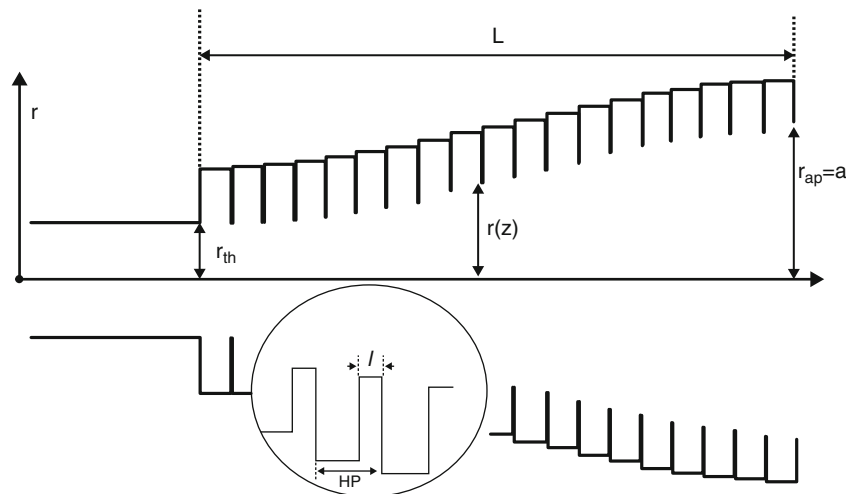
Depending on to what plane its opening tapers rectangular horns may be classified as E-plane sectoral, H-plane sectoral, or pyramidal horns (Figure 5). The type, direction, and amount of taper can have a significant effect on the overall performance of the horn. The radiation characteristics can be determined using the aperture field



Microwave Horn Antennas, Figure 8 Gain characteristics of conical horns of optimum-gain design.



Microwave Horn Antennas, Figure 9 Dual-mode conical Potter horn: (a) horn structure and (b) E-plane pattern.



Microwave Horn Antennas, Figure 10 Profiled conical corrugated horn.

method or GTD construction (Jull, 1973; Huang et al., 1983). The latter is also capable of predicting the near field and back-radiated field. Considerable amounts of design data are available on rectangular horns, and the reader is referred to the just-mentioned references (Balanis, 1988, 1996; Bird and Love, 2007).

For pyramidal horns, the geometrical parameters may be chosen to achieve the so-called optimum-gain design (Balanis, 1988). Figure 6 shows the relationship between the parameters to design an optimum-gain horn for a specified gain (the overall efficiency of these horns is typically about 50 %). Figure 7 is the plot of the E- and H-plane patterns of such horns.

Conical horns

The geometry of a circular horn is shown in Figure 8. In contrast to pyramidal horns which are typically fed by rectangular waveguides, the circular horn is usually fed by circular waveguides. The aperture field of these horns can be constructed in a fashion similar to that of pyramidal horns by simply multiplying the aperture field of the circular waveguide by a quadratic phase term (Balanis, 1988). The resulting integral for the computation of the radiated field can then be evaluated numerically. Figure 8 gives the proper horn dimensions for constructing an optimum-gain horn for a specified gain. These horns typically possess more symmetric E- and H-plane patterns than do their pyramidal horn counterparts. They can also be used more effectively to create circularly polarized fields.

Multimode, corrugated, matched horns, and others

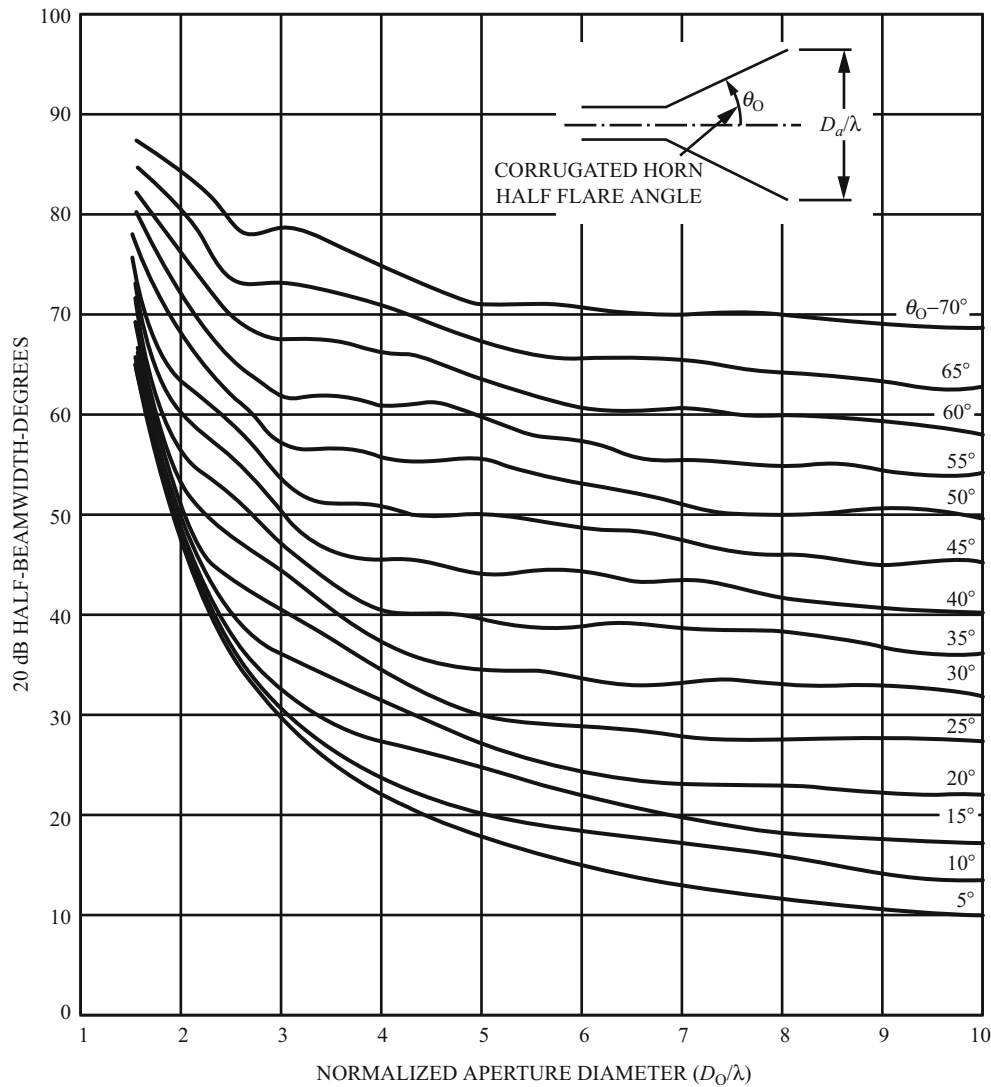
The requirements for designing more advanced remote sensing systems have resulted in the generation of a new class of feeds. These feeds in particular are used in large reflectors for radio astronomy and remote sensing reflectors. Among them one may refer to multimode (Potter, 1963; Thomas, 1970;

Bhattacharyya and Goyette, 2004), corrugated (hybrid mode) (Clarricoats and Olver, 1984; James, 1984; Thomas et al., 1986; Olver et al., 1994), and matched horns (Rudge and Adatia, 1975; Bahadori and Rahmat-Samii, 2006).

Multimode horns were invented to equalize the pattern asymmetry of single-mode horns. For instance, in the Potter horn (Potter, 1963) the TM_{11} mode is generated along with the dominant TE_{11} mode of a circular horn. Although this new TM_{11} mode does not have any appreciable effect on the H-plane radiation pattern, when it is properly phased and combined with the TE_{11} mode, it can effectively alter the E-plane aperture distribution, which results in a symmetric radiation pattern. All these favorable features, however, cannot be properly realized until the feed aperture diameter exceeds about 1.3λ . For example, Figure 9 shows the radiation pattern of a dual-mode Potter horn as reported in Potter (1963). A partial conversion of the TE_{11} mode energy to a TM_{11} mode happens in the flared section of the horn, while the straight section enforces the condition that both modes have the proper phase relation at the aperture which can be maintained over a bandwidth of less than 10 %. There are also available other types of multimode horns, which result from a combination of modes such as TE_{10} , TE_{12} , and TM_{12} in a square-aperture pyramidal horn.

Corrugated horns have become the main choice of antenna feeds in recent advanced communications, radar, and remote sensing applications where demanding performance is required. First introduced in the 1960s (Minnett and Thomas, 1966; Simmons and Kay, 1966), they can provide superior radiation performances such as symmetric patterns and low cross-polarization levels compared with conical horns.

Corrugated horns are capable of creating similar boundary conditions at all polarizations which result in similar tapers in the aperture field distribution in all planes. Due to these boundary conditions, symmetric radiation



Microwave Horn Antennas, Figure 11 20 dB half beamwidths of corrugated horns.

patterns can be obtained at levels as low as -25 dB in both the E- and H-planes. A corrugated horn can be realized by grooving the E-plane of a pyramidal horn or the entire wall of a circular horn with, typically, six or more slots (corrugations) per wavelength (Figure 10). For circular corrugated horns, Figure 11 shows the plots of the pattern widths at 20 dB level as a function of the opening angle.

The existence of the corrugations, especially near the waveguide-horn junction, affects the VSWR of the horn. The usual practice is to begin the corrugations at a small distance from the junction. These horns are also classified as hybrid-mode horns because they support modes in which both longitudinal E- and H-field components are present. In a circular corrugated horn, a natural mixture of TE_{11} and TM_{11} results in the generation of a hybrid-mode HE_{11} . In contrast to dual-mode horns there is no need of a mode converter, and therefore, the

hybrid-mode horns have typically wider bandwidths. In particular, one version of these hybrid horns (scalar horns), which has a large flare angle with a relatively short horn, has radiation characteristics with little dependence on frequency. There now exists advanced computational and optimization methods (Sinton et al., 2002) to tailor corrugated horn antennas for various applications including profiled corrugated horns with shorter lengths than the standard straight corrugated horns.

There are other horn feeds which are properly tailored to specifically overcome some undesirable characteristics of reflectors. For example, matched horn feeds (Rudge and Adatia, 1975; Bahadori and Rahmat-Samii, 2006) are used to significantly reduce the generation of unwanted cross-polarized fields in an offset parabolic reflector illuminated with a tilted horn feed. The basic concept behind these designs is to match the horn

aperture field distribution with the receiving focal plane distribution of the reflector.

There also exists a class of dielectric horns with some unique properties, as described in Oh et al. (1970), Baldwin and McInnes (1973), and Jha et al. (1975), and other configurations (Love, 1962; Vokurka, 1979).

Summary

Microwave horn antennas are widely used as standard gain antennas for antenna testing or evaluation or feeds for other antennas due to their simple and solid geometries and excellent performances. Among many different types of horns, rectangular pyramidal horns and conical horns, multimode and corrugated horns are the most popular and practical configurations, and a selection of a specific type is made based on the design specifications such as bandwidth, gain and aperture efficiency, polarization, and some other unique radiation characteristics. With the advances in computational and optimization methods, more sophisticated designs such as profiled corrugated horns and matched horns are developed for demanding applications in radars, satellite communications, remote sensing, deep-space telemetry, and radio astronomy.

Bibliography

- Bahadori, K., and Rahmat-Samii, Y., 2006. "A Tri-Mode Horn Feed for Gravitationally Balanced Back-to-Back Reflector Antennas", *IEEE Antennas and Propagation Society Symposium*, pp. 4397–4400, Albuquerque, NM.
- Bahadori, K., and Rahmat-Samii, Y., 2009. "Tri-Mode Horn Feeds Revisited: Cross-pol Reduction in Compact Offset Reflector Antennas," *IEEE Trans. Antennas Propag.*, vol. 57, no. 9, pp. 2771–2775.
- Balanis, C. A., 1988. Horn antennas. In Lo, Y. T., and Lee, S. W. (eds.), *Antenna Handbook*. New York: Van Nostrand Reinhold.
- Balanis, C. A., 1996. *Antenna Theory: Analysis and Design*, 2nd edn. New York: Wiley.
- Baldwin, R., and McInnes, P. A., 1973. Radiation patterns of dielectric loaded rectangular horns. *IEEE Transactions on Antennas and Propagation*, **21**, 375–376.
- Bhattacharyya, A., and Goyette, G., 2004. A novel horn radiator with high aperture efficiency and low cross-polarization and applications in arrays and multibeam reflector antennas. *IEEE Transactions on Antennas and Propagation*, **52**, 2850–2859.
- Bird, T. S., and Love, A. W., 2007. Horn antennas. In Volakis, J. (ed.), *Antenna Engineering Handbook*, 4th edn. New York: McGraw-Hill.
- Clarricoats, P. J. B., and Olver, A. D., 1984. *Corrugated Horns and Microwave Antennas*. London: Peregrinus.
- Huang, J., Rahmat-Samii, Y., and Woo, K., 1983. A GTD study of pyramidal horns for offset reflector antenna applications. *IEEE Transactions on Antennas and Propagation*, **31**, 305–309.
- James, G. L., 1984. Design of wide-band compact corrugated horns. *IEEE Transactions on Antennas and Propagation*, **32**, 1134–1138.
- Jha, R. K., Misra, D. K., and Jha, L., 1975. Comparative study of dielectric and metallic horn antennas. In *IEEE Antennas and Propagation Society International Symposium*, pp. 16–18.
- Jull, E. V., 1973. Errors in the predicted gain of pyramidal horns. *IEEE Transactions on Antennas and Propagation*, **21**, 25–31.
- Love, A. W., 1962. The diagonal horn antenna. *Microwave Journal*, **5**, 117–122.
- Love, A. W. (ed.), 1976. *Electromagnetic Horn Antennas*. New York: IEEE Press.
- Minnett, H. C., and Thomas, B. M., 1966. A method of synthesizing radiation patterns with axial symmetry. *IEEE Transactions on Antennas and Propagation*, **14**, 654–656.
- Oh, L. L., Peng, S. Y., and Lunden, C. D., 1970. Effects of dielectrics on the radiation patterns of an electromagnetic horn. *IEEE Transactions on Antennas and Propagation*, **18**, 553–556.
- Olver, A. D., Clarricoats, P. J. B., Kishk, A. A., and Shafai, L., 1994. *Microwave Horns and Feeds*. New York: IEEE Press.
- Potter, P. D., 1963. A new horn antenna with suppressed sidelobes and equal beamwidths. *Microwave Journal*, **6**, 71–78.
- Rudge, A. W., and Adatia, N. A., 1975. New class of primary-feed antennas for use with offset-parabolic-reflector antennas. *Electronics Letters*, **11**, 597–599.
- Simmons, A. J., and Kay, A. F., 1966. The scalar feed – A high performance feed for large paraboloid reflectors. In *IEE Conference Publications*, Vol. 21, pp. 213–217.
- Sinton, S., Robinson, J., and Rahmat-Samii, Y., 2002. Standard and micro genetic algorithm optimization of profiled corrugated horn antennas. *Microwave and Optical Technology Letters*, **35**, 449–453.
- Thomas, B. M. A., 1970. Prime-focus one- and two-hybrid-mode feeds. *Electronics Letters*, **6**, 460–461.
- Thomas, B. M. A., James, G. L., and Greene, K. J., 1986. Design of wide-band corrugated conical horns for Cassegrain antennas. *IEEE Transactions on Antennas and Propagation*, **34**, 750–757.
- Vokurka, V. J., 1979. Elliptical corrugated horn for broadcasting satellite antennas. *Electronics Letters*, **15**, 652–654.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)
[Emerging Technologies, Radar](#)
[Microwave Radiometers](#)
[Microwave Radiometers, Conventional](#)
[Radars](#)

MICROWAVE RADIOMETERS

Niels Skou
 National Space Institute, Technical University of
 Denmark, Lyngby, Denmark

Synonyms

Passive microwave radiometer (PMR)

Definition

The microwave radiometer is a calibrated receiver that measures properties of the natural emission from the environment as picked up by an antenna system.

Introduction: what is radiometry about?

All bodies at a temperature above the absolute zero ($0\text{ K} = -273\text{ }^{\circ}\text{C}$) radiate power, according to Planck's law. At microwave frequencies the Rayleigh-Jeans approximation holds, and the radiation is proportional to physical temperature. Actually, this is only true for the

so-called blackbodies, which are perfect emitters. Natural bodies radiate less, and we introduce the term emissivity (ϵ) which is a number between 0 and 1 describing how well the body radiates relative to a blackbody. Within radiometry the radiated power is expressed as the so-called brightness temperature, T_B , so that $T_B = \epsilon \cdot T_{\text{phys}}$. The brightness temperature of a blackbody is thus equal to its physical temperature, while all natural bodies will have brightness temperatures lower than that. The brightness temperature of a natural body can also be understood as that physical temperature a blackbody would have to have in order to emit the power in question. Typical values related to planet Earth range from almost 0 K (looking up toward free space) to about 300 K.

The radiated energy can be picked up by an antenna in order to be measured by a radiometer. Unfortunately antennas are not perfect: a perfect antenna would have a sharply defined main lobe so that when this is pointed toward the target, only this contributes to the received power. But the main lobe is not with sharp cutoff, so some power is received from the surroundings of the target. Furthermore, antennas have side lobes far from the main lobe meaning that some power is also received from other directions. The process can be expressed by the following equation:

$$T_A = \frac{1}{4\pi} \cdot \int_{4\pi} T_B(\theta, \phi) \cdot G(\theta, \phi) d\Omega. \quad (1)$$

This convolution integral, where $G(\theta, \phi)$ is the antenna radiation pattern, shows how the received power, denoted as the so-called antenna temperature T_A , is a gain-weighted summation of the brightness temperatures $T_B(\theta, \phi)$ from each direction (see Ulaby et al., 1981).

Sensitivity

As stated above, the task of the radiometer is to measure the power picked up by the antenna. So in fact the radiometer is basically a calibrated microwave receiver. The basic and simplest radiometer, the total power radiometer, is illustrated in Figure 1a.

The received power is amplified with a gain G , a certain bandwidth B around the center frequency is selected by the filter, the microwave signal is detected by a square law detector, and since we are handling noise-like signals, a certain integration time τ is required. Finally, it is indicated that we cannot make a receiver without introducing an additional noise T_N .

The output will represent a power measure expressed as $P = k \cdot B \cdot G \cdot (T_A + T_N)$ where k is Boltzmann's constant: $1.38 \cdot 10^{-23}$ J/K. As already stated the signals are noise-like, so the output fluctuates, but the fluctuations are smoothed by the integration. The level of fluctuations, the standard deviation of the output to be specific, is expressed as the radiometer sensitivity, ΔT , and it is calculated from Eq. 2:

$$\Delta T = \frac{T_A + T_N}{\sqrt{B \cdot \tau}}. \quad (2)$$

Stability and accuracy

A fundamental problem with this simple and direct implementation of the radiometer is that the output, $P = k \cdot B \cdot G \cdot (T_A + T_N)$, is dependent on stability of the gain G and the receiver noise temperature T_N . Hence, stability may be problematic, and frequent calibration is required, or a more developed radiometer principle must be employed (Skou and LeVine *cpoo*).

If k , B , G , and T_N are not only constants but known constants, we have no absolute accuracy problems: a given T_A results in a given P that can be calculated. Such knowledge of the constants is rarely available, leading to the necessity for calibration. This illustrates the fundamental difference between stability and accuracy: stability is a highly appreciated virtue of an instrument, but a stable instrument need not be accurate. The steps toward accuracy include the calibration process.

In the following we will describe a slightly different aspect of absolute accuracy, which stresses the care that must be exercised when designing or working with radiometers. Consider losses in a signal path – it could be the waveguide connecting the antenna with the radiometer input or a passive component in the radiometer front end. Let ℓ denote the fractional loss (or the absorption coefficient) and T_0 the physical temperature. T_1 is the input temperature and the output temperature is then given by $T_2 = T_1 (1 - \ell) + \ell \cdot T_0$.

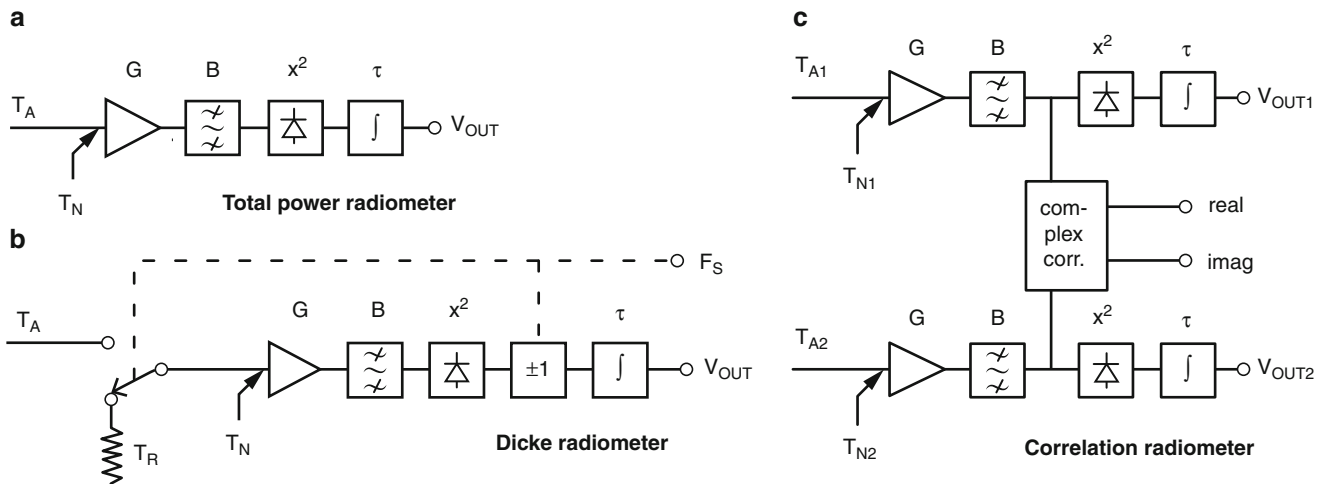
The difference between output and input is $T_D = T_2 - T_1 = \ell (T_0 - T_1)$. If T_1 is 100 K and T_0 is 300 K, a loss as small as 0.01 dB ($\ell = 0.0023$) results in a difference, T_D , of 0.5 K. Bearing in mind that the losses of a real signal path are much greater than 0.01 dB, the physical temperature of the path must be measured and used for correction of the measured brightness temperature. The corresponding losses must be known to an accuracy of better than 0.01 dB and must remain stable within the same limits.

Consider a mismatch, for example, at the input of a radiometer; with a reflection coefficient ρ , we similarly find $T_D = \rho (T_{\text{RAD}} - T_1)$ where T_{RAD} is the microwave temperature as seen from the point of reflection into the radiometer. T_{RAD} is typically 300 K and if T_1 again is assumed to be 100 K, a reflection coefficient of -26 dB will give an error (T_D) of 0.5 K. Care must be exercised to obtain reflection coefficients better than -26 dB.

Radiometer types

The conventional radiometers are illustrated to the left in Figure 1, while the more advanced types are illustrated by the correlation radiometer to the right.

The total power radiometer has already been discussed. It is the simplest and basic radiometer that has the optimum sensitivity but may have problems with stability.



Microwave Radiometers, Figure 1 Two conventional radiometer types: (a) total power and (b) Dicke and (c) the correlation radiometer.

The classical way of alleviating stability problems is to employ switching as done in the Dicke radiometer (Figure 1b). The idea here is that the radiometer in fact measures the difference between the unknown antenna temperature and an internal, known reference. The Dicke radiometer has proven very good behavior and has been widely used. An enhanced version of the Dicke radiometer is the noise injection radiometer (NIR) in which a null-balancing technique is employed resulting in total independence of gains and receiver noise temperatures. A price has to be paid for the enhanced stability of Dicke type switching radiometers: the sensitivity is degraded by a factor of 2 compared with that of the total power radiometer.

More advanced radiometers are very often based on the correlation radiometer shown in Figure 1c. It basically consists of two total power radiometers, but in addition to the usual detected outputs, the signals before detection in the two channels are cross correlated to produce two additional outputs: the real and imaginary part of the correlation. The correlation radiometer is the building block of interferometric, synthetic aperture radiometer systems and often also of polarimetric radiometer systems.

Radiometer antennas

An important part of any radiometer system is its antenna. The purpose of the antenna is to collect the emitted energy from a target and present it to the radiometer input. Thus, the antenna radiation pattern determines the spatial resolution of the radiometer system. Important antenna parameters are main beam efficiency and side lobe level. A good beam efficiency and low side lobes ensure that the radiometer system primarily measures what the antenna points toward and not so much the surroundings. Of paramount importance for a radiometer antenna is low loss following the previous accuracy discussion.

Hence, the most popular antenna types are the microwave horn and reflector antennas.

Horns are bulky, low-gain devices and as such normally not used as primary spaceborne antennas, but they play an important role in ground-based and airborne systems and as feeds for reflector antennas. Especially Potter horns and dual-mode horns are favored due to excellent radiation patterns.

Reflector antennas – typically based on offset parabolic reflectors – have been used extensively in spaceborne systems. They feature low loss and good radiation patterns, and, very importantly, they can serve a wide range of frequencies (e.g., 5–90 GHz) through the use of clusters of feed horns.

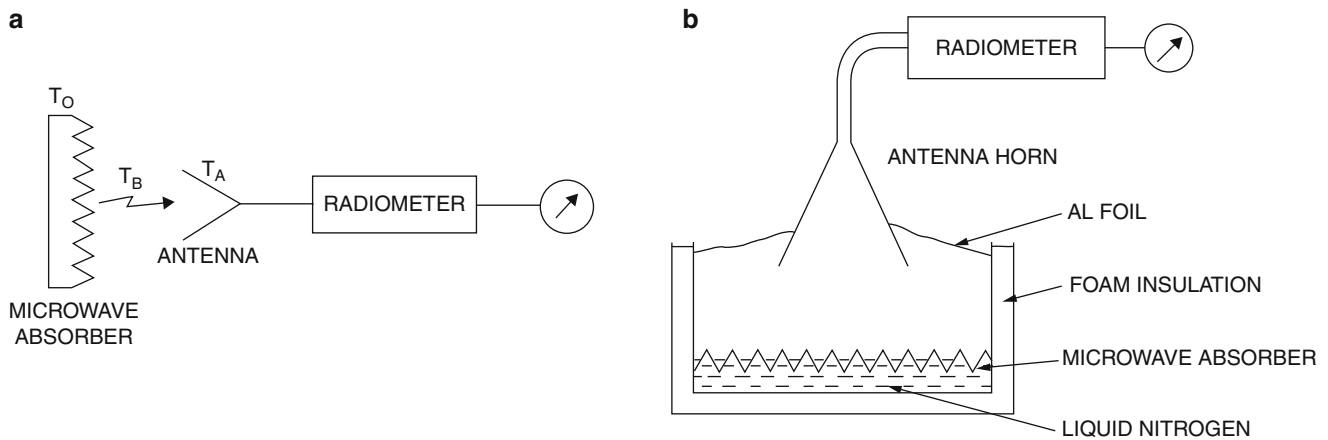
For the antennas discussed here, where the electrical aperture is roughly equal to the physical aperture, it is very simple to estimate the width of the main beam. The 3 dB beamwidth is slightly larger than the reciprocal of the aperture measured in wavelengths of the operating frequency:

$$\theta = 1.4 \cdot \frac{\lambda}{D}. \quad (3)$$

The factor 1.4 reflects the need for high beam efficiency. A 1 m reflector antenna used at 30 GHz (1 cm wavelength) will thus have a 0.014 ($= 0.8^\circ$) beamwidth. Looking straight down from a satellite in an 800 km orbit, the footprint on Earth will be 11.2 km. Spaceborne radiometer systems are not high-resolution devices like optical scanners or synthetic aperture radars, but they do supply unique data for geophysical interpretation with frequent coverage of the globe.

Calibration

The purpose of calibration is to establish the relation between the input brightness temperature and the



Microwave Radiometers, Figure 2 Antenna target calibrator, principle, and practical layout.

radiometer output – usually digital numbers following analog to digital conversion of the radiometer output voltage. Assuming that the radiometer transfer is linear, two calibration points are required, typically one hot around 300 K and one cold in the 0–77 K range. The most obvious solution is a cooled microwave termination. A well-matched load will generate a noise temperature equal to its physical temperature – and the physical temperature is easily measured. If the load is left at room temperature, it provides a hot calibration point, and if it is submerged into liquid nitrogen, it will be a cold load at around 77 K. However, the latter is not quite as easy as it sounds: heat flow into the load must be prevented, and the loss in the input transmission line must be corrected for.

An alternative solution is a cooled target viewed by a suitable antenna connected to the radiometer (see Figure 2). A microwave absorber (normally used to cover the inside walls of radio anechoic chambers) will emit a brightness temperature equal to its physical temperature T_O . Under ideal conditions, the antenna will sense nothing but the brightness temperature from the absorber and $T_A = T_B = T_O$. Figure 2 also shows a practical layout of this concept. The radiometer is connected to an antenna horn through a short waveguide (very low losses!). The horn views a microwave absorber soaked with liquid nitrogen. The absorber and the liquid nitrogen are contained in an insulated metal bucket, and the excess opening of the bucket is covered by aluminum foil. In this way the antenna is only able to pick up energy from the absorber, which is cooled to 77 K by the nitrogen. There is generally no problem with losses and heat flow in the antenna and the waveguide, liquid nitrogen is readily available, and the setup is cheap and simple. Overall this is a very useful radiometer calibrator – and before liquid nitrogen is applied it provides also the hot calibration point.

Finally, it shall be noted that pointing an antenna toward free space will provide a cold calibration point of only a few K. The problem is that the atmosphere will give

a significant contribution to the temperature, and it must be carefully accounted for. However, this is not the case for satellite-borne systems, and here such a sky view is often used for the cold calibration point.

Going to the more advanced correlation radiometers, also shown in Figure 1, calibration is somewhat more complicated. First, a basic calibration of the two channels is carried out as just described, but in addition to this, the calibration of the correlation channels requires generation of a pair of known signals with a known amount of correlation between them. This issue is outside the scope of the present text.

Conclusion

Microwave radiometers are sensitive receivers requiring special attention to stability and accuracy. Several types have been developed over the years each with its special virtue. The antenna is an important part of the radiometer system determining what is actually measured by the radiometer. Any radiometer must be carefully calibrated – not only after its development but also a frequent check of calibration must be carried out at proper intervals when the radiometer is in use. More information can be found in (Skou and LeVine, 2006).

Bibliography

- Skou, N., and LeVine, D., 2006. *Microwave Radiometer Systems, Design and Analysis*. Boston: Artech House.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing*. Norwood: Artech House, Vol. 1.

Cross-references

- [Calibration, Microwave Radiometers](#)
- [Microwave Horn Antennas](#)
- [Microwave Radiometers](#)
- [Microwave Radiometers, Conventional](#)
- [Microwave Radiometers, Correlation](#)
- [Microwave Radiometers, Interferometers](#)
- [Microwave Radiometers, Polarimeters](#)

MICROWAVE RADIOMETERS, CONVENTIONAL

Niels Skou

National Space Institute, Technical University of
Denmark, Lyngby, Denmark

Synonyms

Passive microwave radiometer (PMR)

Definition

The conventional microwave radiometer is a calibrated receiver that measures the power from the environment as picked up by an antenna system.

Introduction

The emission from the environment is received by the antenna. This can typically be horizontally or vertically polarized. The conventional radiometer measures the power corresponding to one of these polarizations, and it is basically a single channel receiver. If both polarizations must be measured, two independent radiometer receivers are employed. In-depth treatment of conventional radiometers is found in Ulaby et al. (1981) and in Skou and LeVine (2006).

Total power radiometer

The simplest and most direct implementation of a radiometer is found in the total power radiometer (TPR) as illustrated in Figure 1a. The gain in the radiometer has been symbolized by an amplifier with a gain G and the frequency selectivity by a filter with a bandwidth B (centered around some given frequency). We cannot make a receiver without introducing an additional noise T_N . The microwave power has to be detected to find some measure of its mean. In the present case, it is very attractive to use a square-law detector: then the output *voltage* will be proportional to the input *power* and hence the input temperature. Finally, we indicate where the integration takes place: the signal from the detector is smoothed by the integrator to reduce fluctuations in the output, and the longer is the integration time, the more smoothing.

The output can be expressed as

$$V_{\text{OUT}} = c \cdot (T_A + T_N) \cdot G \quad (1)$$

where c is a constant. V_{OUT} is totally dependent on T_N and G . These can for some applications not be regarded as stable enough to satisfy reasonable requirements to stability and accuracy. In other cases, however, the total power radiometer is very useful, namely, where frequent calibration, for example, once every few seconds, is possible.

The sensitivity (standard deviation of the output signal) of the total power radiometer is given by

$$\Delta T_{\text{TPR}} = \frac{T_A + T_N}{\sqrt{B \cdot \tau}} \quad (2)$$

Dicke radiometer

In 1946, R. H. Dicke found a way of alleviating the stability problems in radiometers (Dicke, 1946). By using the radiometer not to measure directly the antenna temperature T_A but rather the difference between this and some known reference temperature T_R , the sensitivity of the measurement to gain and noise temperature instabilities is greatly reduced (see Figure 1b).

The input of the radiometer is rapidly switched between the antenna temperature and the reference temperature. The switch frequency F_S could be 1,000 Hz. The output of the square-law detector is multiplied by +1 or -1, depending on the position of the Dicke switch, before integration. The input to the integrator is then $V_1 = c \cdot (T_A + T_N) \cdot G$ in one half period of F_S and $V_2 = -c \cdot (T_R + T_N) \cdot G$ in the second half period.

Provided that the switch frequency F_S is so rapid that T_A , T_N , and G can be regarded as constants over the period and that the period is much shorter than the integration time, the output of the radiometer is found as

$$V_{\text{OUT}} = c \cdot (T_A - T_R) \cdot G \quad (3)$$

It is seen that T_N has been eliminated, while G is still present, although with less weight. Now G multiplies the difference between T_A and T_R , where T_R is reasonably chosen to be in the same range as T_A , while in the total power case, G multiplied the sum of T_A and the rather large T_N .

The Dicke principle has proven to be very useful, and Dicke radiometers (DR) have been used extensively over the years.

A price has to be paid, however, for the better immunity to instabilities. Since only half of the measurement time is spent on the antenna signal (the other half is spent on the reference temperature), the sensitivity is degraded by a factor of 2 compared with the total power radiometer:

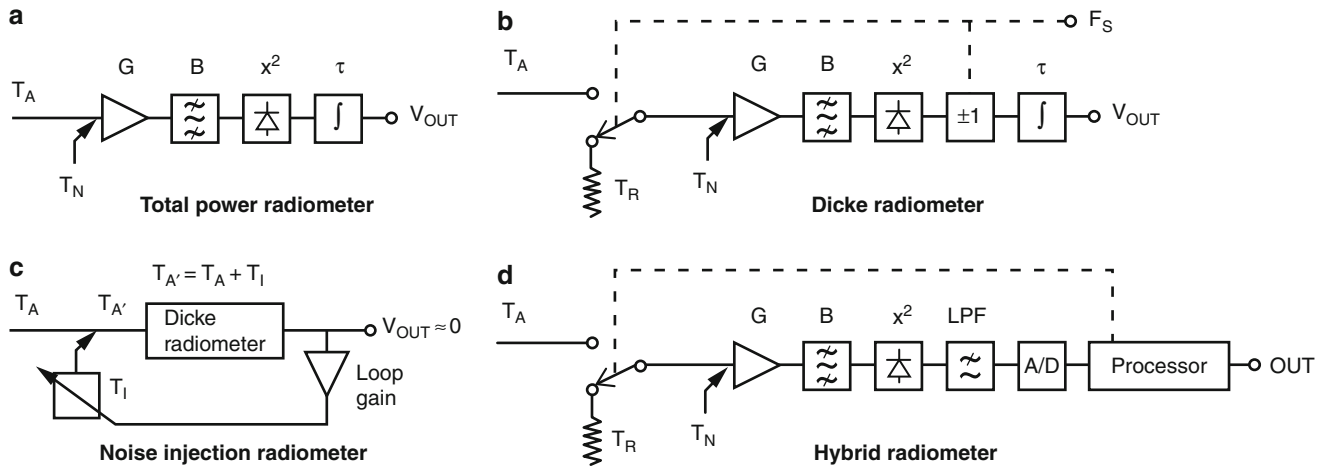
$$\Delta T_{\text{DR}} = 2 \cdot \frac{T_A + T_N}{\sqrt{B \cdot \tau}} \quad (4)$$

Noise injection radiometer

The noise injection radiometer (NIR) represents the final step toward stability; that is, the output is independent of gain and noise temperature fluctuations (Goggins, 1967; Hardy et al., 1974).

From Equation 3 it is seen that the output from a Dicke radiometer is zero (independent of G and T_N) if the reference temperature and the antenna temperature are equal. The noise injection radiometer is a specialization of a Dicke radiometer in which this condition is continuously fulfilled by a servo loop.

In almost any case encountered in Earth remote sensing, the antenna temperature is below some 300 K (emissivities between 0 and 1 are multiplied by the physical temperature). The reference temperature in a Dicke radiometer is conveniently equal to the physical



Microwave Radiometers, Conventional, Figure 1 (a) Total power radiometer, (b) Dicke radiometer, (c) noise injection radiometer, and (d) hybrid radiometer.

temperature in the microwave front end, that is, 300–320 K. In [Figure 1c](#) we show how the output T_I of a variable noise generator is added to the antenna signal T_A , so that the resultant input (T_A') to the Dicke radiometer is equal to the reference temperature (T_R) and a zero output results. A servo loop adjusts T_I to maintain the zero output condition or rather the near-zero output condition: the loop gain can be made large but not infinite.

From [Equation 3](#) we have

$$V_{OUT} = c \cdot (T_A' - T_R) \cdot G = 0$$

and as $T_A' = T_A + T_I$, we find $T_A = T_R - T_I$. T_R is a known constant, and knowledge of T_I is required to find T_A . The accuracies of the Dicke radiometer part of the NIR and of the loop gain are (given large loop gain) completely insignificant for the accuracy with which we determine T_A . This is solely dependent on the accuracy of T_I . Accurate and stable noise sources with variable output can be made and are used for “injecting” the required signal T_I into the input line, so that T_I and T_A are added.

The sensitivity of the noise injection radiometer is very close to that of the Dicke radiometer:

$$\Delta T_{NIR} = 2 \cdot \frac{T_R + T_N}{\sqrt{B \cdot \tau}} \quad (5)$$

Hybrid radiometer

The radiometers as described hitherto in this entry are the classical receiver types, and their implementation is indicated in the classical way using, for example, analog integration after detection and analog subtraction of antenna and reference signals (in the Dicke radiometer). And indeed many radiometers are still implemented this way. But with the advent of analog to digital converters and digital processing, other implementation forms are possible and often used. This is illustrated in [Figure 1d](#).

As soon as possible following detection, the signal is A to D converted – only a low-pass filter is indicated to condition the signal bandwidth to the sampling frequency of the converter. The signal from the converter is led to some kind of digital processor, typically a PC or an FPGA (field-programmable gate array), where suitable data handling takes place. This can typically be digital integration to the required integration time τ , as well as subtraction of the antenna signal and the reference signal. Since these processes are under computer control, flexibility becomes a key word, and the distinction between total power and Dicke radiometer vanishes to some extent: if the processor operates the input switch rapidly and regularly, we can regard it as a Dicke case, while if the measurement situation is such that the antenna signal can be measured (with interruptions when the switch points to the reference signal) without loss of data, then we have a total power case with frequent calibration.

A classical Dicke radiometer spends half of its time measuring the well-known reference temperature and thus only half of its time doing its real job, namely, measure the unknown antenna temperature. Thus over the years researchers have considered a better duty cycle for the antenna measurements (in turn potentially leading to improved sensitivity). However, having analog subtraction of the signals, the 50 % duty cycle is instrumental to having simple and stable circuitry. But with the subtraction done digitally this is no longer a limitation, and optimized duty cycles can be found on a case-by-case basis. A word of warning should, however, be stated: the naive notion that spending more time on the unknown antenna signal will lead to improved sensitivity is not necessarily true! When more time is spent on the antenna, less time is left for the reference and this in turn leads to an increased standard deviation for that measurement. Thus, the final standard deviation after subtraction might not decrease.

However, there are possible improvement schemes: since the reference temperature and the receiver noise temperature are assumed relatively stable while the antenna temperature may change rapidly, averaging over several reference temperature measurements will reduce the standard deviation of this measurement and thus allow a non-50 % duty cycle to be employed. Initial instrument fluctuations and averaging times must of course be considered carefully. See further discussion in Tanner et al. (2003) and Racette and Lang (2005).

Implementation

The radiometer is merely a very sensitive microwave receiver, and like any receiver, it can be implemented as a direct receiver or, by use of a mixer and a local oscillator, as a superheterodyne receiver. In the direct receiver, all amplification takes place at the input RF frequency, and all selectivity is determined by filters in the same frequency range. In the superheterodyne receiver most of the amplification takes place at the much lower intermediate frequency (IF), and selectivity is determined by a combination of filters at RF and IF levels.

Microwave FET amplifiers with excellent noise figures and square-law detectors covering the frequencies well beyond 40 GHz are available. Hence, the general trend is that conventional radiometers be implemented as direct receivers at frequencies below some 40 GHz, while higher-frequency implementations seem to favor the superheterodyne receiver – due to unavailability of amplifiers or very expensive amplifiers.

Another very important issue when implementing microwave radiometers is the concept of thermal stabilization. The best and simplest way to achieve a stable

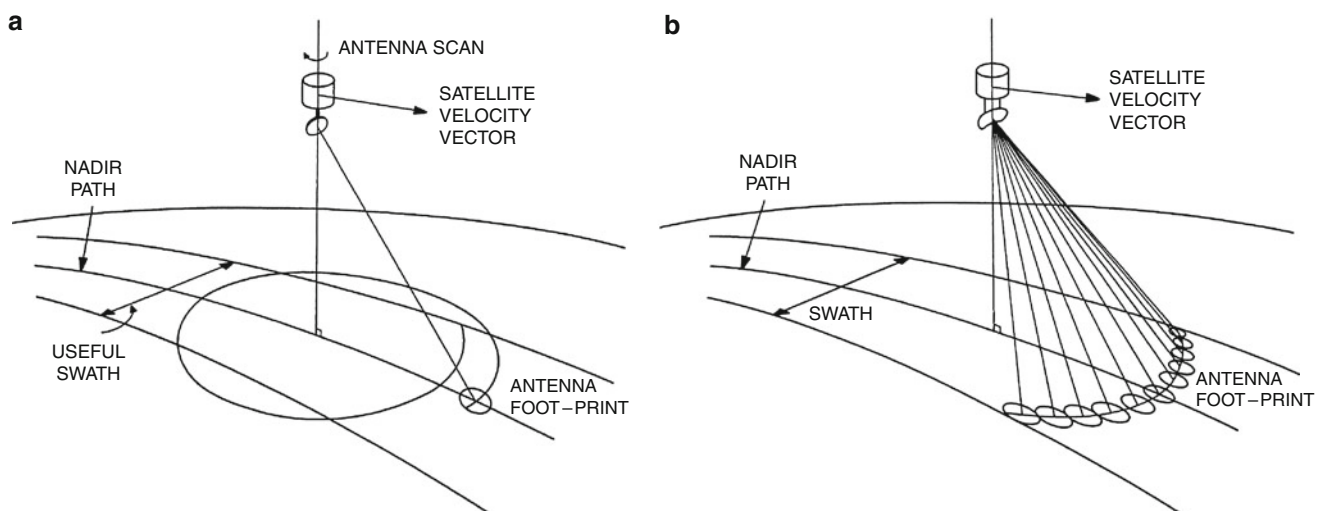
radiometer is to carefully temperature-stabilize all important components of the receiver – first and foremost the microwave front end.

Imaging considerations

A microwave radiometer is often mounted on an aircraft or a satellite in order to cover large areas.

If a radiometer with its antenna is mounted on a satellite and the antenna points straight down, the forward movement of the satellite will facilitate measurements on the ground along a straight line (the nadir path of the satellite). Coverage of the entire Earth by such “profiles” will require an enormous number of orbits! A dramatic increase in mapping efficiency results from scanning the antenna (see Figure 2a). The antenna rotates about a vertical axis, and the footprint will cover a wide swath on the Earth dependent on satellite altitude and incidence angle. Other scanning methods are possible, but the rotating scan is attractive due to constant incidence angle on the ground and the lack of accelerations associated with reciprocating scans.

In a scanning system it is obvious that there is only a limited time for the radiometer to carry out its measurement before the footprint moves to another position within the swath. The so-called dwell time per footprint is short. The users in general want small footprints (or high spatial resolution, to put it differently). As technology evolves, high-resolution systems become possible, but a small footprint results in rapid rotation (mechanical problems) and in a very small dwell time per footprint, hence in a short integration time, which, through our radiometer sensitivity formula, directly translate into poor sensitivity! The solution to this fundamental problem is offered by the so-called push-broom concept illustrated in Figure 2b.



Microwave Radiometers, Conventional, Figure 2 (a) Conical scan and (b) push broom.

In the push-broom radiometer system a multiple-beam antenna covers the swath while the satellite moves forward. A host of radiometer receivers are connected to an equal number of antenna feeds, producing individual beams to sense the Earth simultaneously.

The obvious advantages of the push-broom system compared to a scanning system are no moving antenna to cause problems in the satellite design and much larger dwell time per footprint, hence better sensitivity.

For the scanning radiometer systems, the requirement for receiver sensitivity is severe. At the same time, frequent calibration is easily achieved: once per scan, while the antenna is anyway looking away from the swath, the receiver is calibrated. Hence, the total power radiometer is an obvious candidate for such systems, due to its optimal sensitivity and since potential instabilities are taken care of by frequent calibration.

For the push-broom system, requirements for receiver sensitivity are greatly relaxed due to the much larger dwell time per footprint as compared to the scanner situation. At the same time, frequent calibration is not attractive, as all receivers are always busy sensing the Earth. Hence, the push-broom situation favors a trading of sensitivity for stability, and in conclusion, a Dicke type of switching radiometer is preferred – maybe the NIR.

Conclusion

Microwave radiometers are sensitive receivers requiring special attention to stability and accuracy. Several types have been developed over the years, and in general stability comes at a price: degraded sensitivity. So, when designing a satellite-borne imaging system, an important trade-off between sensitivity, stability, and imaging properties must be carried out.

Bibliography

- Dicke, R. H., 1946. The measurement of thermal radiation at microwave frequencies. *The Review of Scientific Instruments*, **17**, 268–279.
- Goggins, W. B., 1967. A microwave feedback radiometer. *IEEE-AES*, **3**, 83–90.
- Hardy, W. N., Gray, K. W., and Love, A. W., 1974. An S-band radiometer design with high absolute precision. *IEEE-MTT*, **22**, 382–390.
- Racette, P. E., and Lang, R. H., 2005. Radiometer design analysis based upon measurement uncertainty. *Radio Science*, **40**, 1–22.
- Skou, N., and LeVine, D., 2006. *Microwave Radiometer Systems, Design and Analysis*. Norwood: Artech House.
- Tanner, A. B., Wilson, W. J., and Pellerano, F. A., 2003. Development of a high-stability L-band radiometer for ocean salinity measurements. In *IEEE, Proceedings of IGARSS'03*, pp. 1238–1240.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing*. Norwood: Artech House, Vol. 1.

Cross-references

[Calibration, Microwave Radiometers](#)
[Microwave Horn Antennas](#)
[Microwave Radiometers](#)

[Microwave Radiometers, Correlation](#)
[Microwave Radiometers, Interferometers](#)
[Microwave Radiometers, Polarimeters](#)
[Reflector Antennas](#)

MICROWAVE RADIOMETERS, CORRELATION

Christopher Ruf

Department of Atmospheric, Oceanic and Space Sciences,
 University of Michigan, Ann Arbor, MI, USA

Definition and overview

A microwave radiometer measures statistical properties of the electrical field incident on its antenna. A total power radiometer, for example, measures the variance of the field strength, which is proportional to the brightness temperature of the source of the electric field. A correlation radiometer measures the statistical covariance between two incident electric fields. The most common types of correlation radiometers are autocorrelation spectrometers (Ruf and Swift, 1988), spatial interferometers (Kerr et al., 2000), and coherent detection polarimeters (Piepmeier and Gasiewski, 2001). Autocorrelation spectrometers measure the correlation between one electric field and a time-delayed version of itself. Spatial interferometers measure the correlation between the electric field at two locations. Polarimeters measure the correlation between two polarization components of an electric field. Since variance is defined as the statistical correlation of a signal with itself, a total power radiometer can be thought of as a special case of all three correlation radiometers – a spectrometer with zero time delay, an interferometer with no spatial separation, or a polarimeter with a common polarization component.

Autocorrelation spectrometer

The autocorrelation of a time-varying electric field is given by

$$R(\tau) = \langle E^*(t, \vec{r}) E(t - \tau, \vec{r}) \rangle \quad (1)$$

where $E(t, \vec{r})$ is the electric field at time t and position $\vec{r}$, $*$ denotes complex conjugation, τ is the time delay between the two versions of the electric field being correlated, and $\langle \bullet \rangle$ denotes a statistical expectation operator. In practice, the expectation operation is approximated by a time average because, in most cases, second-order statistics of the electric field associated with microwave thermal emission can be considered to be stationary over short time intervals. The autocorrelation is typically sampled over a suitable range of time delays, τ . The Fourier transform of the autocorrelation with respect to the time delay is its power spectrum. For this reason, the autocorrelation of a radiometric signal is often measured in order to determine the spectral dependence of its brightness temperature. The Fourier transform that recovers the power

spectrum from the autocorrelation is performed in software as part of data post-processing. Autocorrelators are used to determine the Tb spectrum when the spectrum is to be resolved with sufficiently fine resolution that a bank of narrow band filters is impractical.

Spatial interferometer

The correlation between the electric field at two positions is given by

$$V(\vec{s}) = \langle A_1 \{E(\vec{r}, \Omega)\}^* A_2 \{E(\vec{r} - \vec{s}, \Omega)\} \rangle \quad (2)$$

where $E(\vec{r}, \Omega)$ is the electric field at position $\vec{r}$ arriving from angular direction Ω , $A_n\{\bullet\}$ denotes the angular reception sensitivity of antenna n ($n = 1, 2$) to the incident electric field, and $\vec{s}$ is the separation between the two versions of the electric field being correlated. This correlation statistic is referred to as the visibility of the brightness temperature distribution. The visibility is sampled over a suitable range of separations between antenna pairs. If the two antennas have overlapping angular reception sensitivities, then the Fourier transform of the visibility with respect to the separation, $\vec{s}$, is related to the angular dependence of the power density of the incident electric field received by the antennas. The power density is, in turn, proportional to the brightness temperature of the source of the electric field. Spatial interferometers (also called Fourier synthesis imagers) are used to determine the angular dependence of the brightness temperature, $T_b(\Omega)$. The Fourier transform required to convert measured visibilities to a Tb image is performed in software as part of the data post-processing. Interferometers are used to image the Tb when its angular variation is to be resolved with sufficiently fine resolution that a large antenna, capable of comparable spatial resolution, is impractical.

Coherent detection polarimeter

The correlation between two linearly polarized components of the electric field is given by

$$C_{pg} = \langle E_p(t, \vec{r})^* E_q(t, \vec{r}) \rangle \quad (3)$$

where the subscripts p and q denote the two polarization components being correlated. A common polarization component was assumed in [Equations 1 and 2](#) for the two versions of the electric field being correlated. If p and q are the same, then C_{pg} is proportional to that polarization component of the brightness temperature of the source of the electric field. If p and q are orthogonal, then the real and imaginary components of C_{pg} are proportional to the third and fourth Stokes parameters in brightness temperature. This type of correlation radiometer differs from the other two in that the correlation measurement itself, and not a Fourier transform of it, is typically the measurement of fundamental interest.

Conclusions

Correlation microwave radiometers measure the statistical covariance between different components (in time, space, or polarization) of the electric field radiated by a source of thermal emission. Certain properties of the brightness temperature associated with that source (its spectrum, its angular distribution, or its polarization state) can be determined from the correlation measurements. Correlation radiometers are typically used when determination of those properties would be otherwise difficult or impossible.

Bibliography

- Kerr, Y., Font, J., Waldteufel, P., and Berger, M., 2000. The soil moisture ocean salinity mission (SMOS). *Earth Observation Quarterly*, **66**, 18.
- Piepmeyer, J. R., and Gasiewski, A. J., 2001. Digital correlation microwave polarimetry: analysis and demonstration. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(11), 2392.
- Ruf, C. S., and Swift, C. T., 1988. Atmospheric profiling of water vapor density with a 205–235 GHz autocorrelation radiometer. *Journal of Atmospheric and Oceanic Technology*, **5**(4), 539–546.

Cross-references

[Microwave Radiometers, Interferometers](#)
[Microwave Radiometers, Polarimeters](#)

MICROWAVE RADIOMETERS, INTERFEROMETERS

Manuel Martin-Neira

European Space Agency (ESA-ESTEC), Noordwijk,
The Netherlands

Synonyms

Aperture synthesis radiometers; Synthetic, interferometric radiometers

Definition

Interferometry. Technique of using the mean product of two random signals to infer some characteristics of the source that generated them.

Aperture synthesis. Application of interferometry to an array of antennas to form an image of the signal source equivalent to one that would be observed by an antenna the size of the whole array.

Introduction

Microwave interferometric radiometers are a particular class of microwave radiometers, which work on the same principles as radio telescopes (Krauss, 1996; Thompson et al., 1988): interferometry and aperture synthesis. In turn, the development of the first aperture synthesis radiometers for space applications has led to a fundamental review of the very principles of radio astronomy

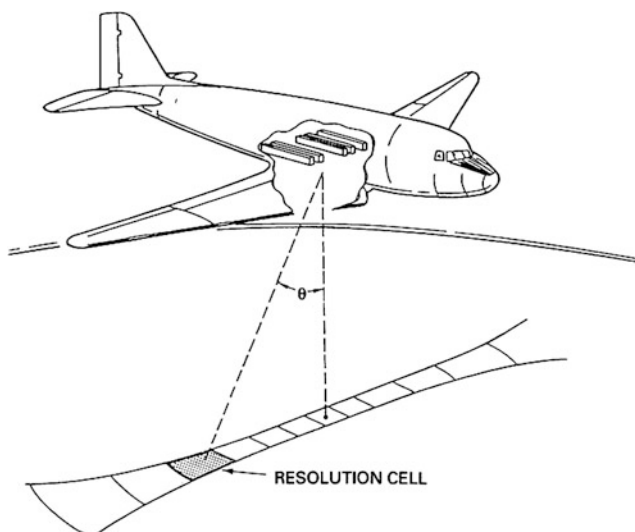
exemplified in the *Corbella equation*. Instead of being pointed to the sky, these radiometers observed the Earth surface from ground-based, airborne, and spaceborne platforms. However, interferometric radiometers present very different features by comparison with radio telescopes, namely, the small size of the antenna elements and the short spacing at which they are clustered together, both of the order of the wavelength. Aperture synthesis radiometers have been developed to achieve fine spatial resolution in those applications where a single scanning antenna is mechanically too complex to realize. Since an interferometric radiometer does not need to scan to make an image, this has been the preferred solution in those cases. The original driving scientific applications have been the mapping of soil moisture and ocean salinity from space (Swift, 1993) and, more recently, atmospheric observations from geostationary orbit.

Historical development of aperture synthesis in remote sensing

While the origin of radio astronomy dates back to 1950 (Ryle et al., 1950), its application to Earth observation was only considered in the late 1970s by the University of Berne (Schanda, 1979) and the beginning of the 1980s by engineers at NASA Goddard Space Flight Center in collaboration with the University of Massachusetts at Amherst (LeVine and Good 1983; Ruf et al., 1988; Tanner, 1990; Swift et al., 1991). The objective behind was to map the Earth's soil moisture and ocean salinity, two important geophysical parameters never measured before at global scale.

The first interferometric radiometer that was built had a synthetic beam in only one dimension, using the real aperture antenna pattern in the other. This was NASA's ESTAR (electronically steered thinned array radiometer), an aircraft demonstrator of such a hybrid instrument; see Figure 1 (LeVine et al., 1992; Swift, 1993). Subsequent developments followed elsewhere with different variations as that using the motion of the platform to save in the required number of receivers (Komiya, 1990), this being equivalent to the use of Earth rotation in radio astronomy.

Aperture synthesis in two dimensions was developed in Europe during the 1990s. The Technical University of Denmark constructed a laboratory demonstrator (Laursen et al., 1994; Skou, 2004), and the European Space Agency (ESA) started the research of an L-band spaceborne MIRAS (Microwave Imaging Radiometer with Aperture Synthesis) (Martin-Neira, 1993; Goutoule et al., 1994; Bayle et al., 2002). ESA's study involved French scientists at the Centre d'Etudes Spatiales de la Biosphère (CESBIO) (Kerr et al., 2000) and benefited with the participation of radio astronomers from the Observatoire du Midi Pyrenees (Lannes and Anterrieu 1994; Anterrieu et al., 2002). The Polytechnic University of Catalonia (UPC) in Barcelona played an important role in defining the requirements and calibration strategy for MIRAS



Microwave Radiometers, Interferometers, Figure 1 NASA's electronically steered thinned array radiometer: the first one-dimensional airborne radiometer (Courtesy David LeVine, NASA).



Microwave Radiometers, Interferometers, Figure 2 Helsinki University of Technology's two-dimensional airborne interferometer (HUT-2D) (Courtesy Martti Hallikainen, TKK).

(Camps, 1996; Torre et al., 1996). Even more crucial was UPC's research on the completed ESA's MIRAS demonstrator, which led to the *Corbella equation* (Corbella et al., 2004), a fundamental correction to the formulation used by radio astronomers. The Helsinki University of Technology embarked in the manufacturing of an airborne two-dimensional microwave interferometer, the HUT-2D (see Figure 2) (Rautiainen et al., 1999), the first airborne two-dimensional aperture synthesis radiometer that provided good quality images of the Earth surface (Kainulainen et al., 2007). The calibration strategy of SMOS was first tested on HUT-2D (Colliander et al., 2007).



Microwave Radiometers, Interferometers, Figure 3 ESA's soil moisture and ocean salinity (SMOS) mission (Courtesy Yann Kerr, CESBIO).

In 1999, the SMOS (soil moisture and ocean salinity) mission was selected by ESA as second Earth Explorer Opportunity Mission, carrying MIRAS as only payload (McMullan et al., 2008). SMOS was launched 2 November 2009 being the first aperture synthesis radiometer flown in space; see Figure 3. It has successfully demonstrated the technique paving the way for its application in other areas. In fact, aperture synthesis has been proposed from geostationary orbit (Ruf, 1990), and higher frequency interferometers are now considered viable for Earth observation satellites flying in low Earth orbit.

In the 2000s, several other ground-based and airborne microwave interferometric radiometers have been developed by different groups around the world, as NASA Goddard's ESTAR-2D, JPL's and ESA's geostationary sounder demonstrators (Christensen et al., 2007), the C- and X-band one-dimensional interferometers of the Chinese Center of Space Science and Application Research's (Wu et al., 2005; Yan et al., 2005), or ESA's Airborne MIRAS in Europe.

Basic principles in remote sensing aperture synthesis

A microwave aperture synthesis radiometer consists of a collection of N antennas arranged in an adequate geometry for optimum imaging. Every antenna is connected to a microwave receiver, which filters out all frequencies of the incoming radiation except for a narrow band, in such a way that the output signal is quasi-monochromatic, with slowly varying but random amplitude and phase. The receiver output signal is then a random process with *Gaussian statistics* and is best represented by its complex *analytic signal*.

Each of the $N \times (N - 1)/2$ possible pairs that can be formed with all the antennas of the collection is called a *baseline*. The two elements of a baseline receive the radiation from a distant target with some delay with respect to each other depending on the angle of arrival of the signal relative to the baseline. The average product of the two receiver output signals of every baseline, named *correlation* or visibility, is then computed. The visibility presents maxima and minima (null value), depending on whether the corresponding spatial delay is or not an integer number of the central wavelength. This means that some directions yield peak visibility values, while others do not contribute to it at all, what is equivalent to a spatial filtering. Every baseline acts then as a spatial filter: the longer the baseline, the shorter the spatial wavelength of the filter. Baselines oriented in different directions provide spatial filtering along those same directions. Thus, the set of all visibilities provides the spatial frequency content of the scene inside a frequency domain limited by the physical extent and location of the antenna elements. The image can then be recovered by a Fourier Transform of the visibilities. This is the formulation of the Van Cittert–Zernike theorem. Nonetheless, as explained later, this theorem would not suffice to describe the operation of an aperture synthesis radiometer, which obeys the Corbella equation instead.

Spatial resolution and windowing

By the properties of the Fourier Transform, the angular resolution of the interferometric radiometer is determined by the extension of the spatial frequency domain. Since a given baseline can be taken in one direction as well as in its opposite direction, the spatial frequency support has a size twice the maximum baseline length, this fixing the angular resolution.

The angular resolution of an aperture synthesis radiometer can now be compared with that of a real aperture instrument. An antenna of the same physical size as the synthetic radiometer would produce a far-field distribution equal to the Fourier Transform of the aperture illumination, following standard antenna theory. The square of the far field gives the radiated power that, when inverse Fourier Transformed, yields the spatial convolution of the original field illumination with itself. For a uniformly illuminated square aperture, the self-spatial convolution becomes a square pyramid with a side twice the physical length of the aperture. This is equivalent to applying a pyramidal weighting function to the visibilities collected by the synthetic radiometer.

In summary, the same spatial resolution of a real aperture with a given illumination can be achieved by an interferometric radiometer of the same physical size and a proper weighting of its visibility function. This is the principle of *aperture synthesis* or *aperture thinning*.

Real aperture radiometers never use uniform illumination as the side lobes become too high. Instead strong tapering is applied to increase the amount of energy collected through the main beam. Similarly, the visibility

samples of an aperture synthesis radiometer are multiplied by weighting functions, or *windows*, that achieve a better beam efficiency than when using a pyramidal weighting. As an example, SMOS weighs the visibility function with a *Blackman window*.

Field of view

The signal received by each element of an interferometer is affected by its antenna pattern. As the gain of the element decreases substantially outside its main beam, the radiometer becomes quite insensitive away from boresight. Imaging is therefore only feasible within the main beam of the element, which defines the field of view of the interferometric radiometer. Small antenna elements are preferred because of their wide beam, which avoids the need for mechanical scanning, a major advantage of interferometric radiometers. This is an important difference with respect to radio astronomy, where mechanically scanned large dish antennas are used. The field of view of an aperture synthesis radiometer is hence very wide and constrained by the element pattern, but not only, as explained below.

Spatial frequency sampling and aliases

The element pattern of a microwave interferometer is chosen to fit the size of the area to be imaged, or *swath*, in each particular application. For instance, an interferometer flying in a low Earth orbit at 800 km altitude will map the complete Earth in less than 3 days if its field of view is 60° wide (900 km on ground), while from geostationary orbit 18° suffices to cover the Earth disk. In matching the element pattern to the area of interest, the physical size of the element becomes defined, which in turn determines the pitch or minimum *spacing* possible between a pair of adjacent antenna elements. The resulting element spacing is usually larger than the maximum sampling period of the spatial frequency domain that guarantees the absence of grating lobes, and *aliases* appear in the image surrounding the alias-free area. The alias borders, like the element pattern, limit the field of view in aperture synthesis radiometry.

The extension of the alias zones depends on the physical arrangement of the antenna elements. For any given spacing, the alias is minimized when the antennas are placed on a hexagonal grid with a pitch the same as the spacing. Examples of such hexagonal geometries are the triangle, the hexagon, the Y (three lines at 120°), and, in general, any snowflake configuration, which lead to visibility domains with the same shapes (triangle, hexagon, a six-point star). An alternative geometry is the rectangular one, like a square, a rectangle, a cross, a U, or a T, all leading to a square or rectangular coverage in the spatial frequency domain. The maximum distances between elements (normalized to the wavelength) before aliases appear are 0.577 and 0.5 for the hexagonal and rectangular geometries, respectively, which shows the benefit of the former.

Radiometric sensitivity

The sensitivity of an aperture synthesis radiometer is comparable to that of a real aperture radiometer. As an example, consider a U-shape interferometer with N elements per arm. On the one hand, the sensitivity is proportional to the collecting area, given by the area of each antenna element a times the number of them $3N$, that is, $3Na$. On the other, the sensitivity is proportional to the square root of the integration time, or, what is equivalent, proportional to the square root of the number of resolution cells inside the field of view. In alias-free conditions, this number equals $2N$. Finally, the sensitivity is also proportional to the volume of the weighting window, normalized by the area of its base, which for a pyramidal window gives $1/3$. The final sensitivity of the interferometer is thus determined by $2N^2a$. A uniformly illuminated real aperture radiometer of collecting area A has 4 resolution cells in its field of view, and its sensitivity is thus proportional to $2A$. If both instruments have the same physical size, $A = N^2a$, the sensitivity of the real aperture radiometer becomes $2N^2a$, the same as that of the aperture synthesis radiometer.

The Corbella equation

The Van Cittert–Zernike theorem, basis of radio astronomy, is not compatible with the Bosma theorem (Wedge and Rutledge 1991). This became apparent for the first time within ESA's MIRAS Demonstrator Pilot Project in 2002. A two-dimensional aperture synthesis radiometer was placed inside an anechoic chamber, and all measured visibilities were zero, as expected from the Bosma theorem. This theorem states that for any passive network (as an array of antennas inside an anechoic chamber) in thermal equilibrium with its terminations (as the isolators following the antennas), the cross-correlation of the outgoing noise waves must be null. But according to the Van Cittert–Zernike theorem, the measured correlations should have been equal to the Fourier Transform of the element antenna pattern. The theoretical foundations of aperture synthesis were revised and a new equation established, compatible with the Bosma theorem, which correctly explains aperture synthesis, and by extension, radio astronomy: the *Corbella equation*. According to this equation, the visibility function is proportional to the difference or *contrast* in physical temperature between the target and the instrument. High correlations are expected when looking to the cold sky, and very small correlations when inside an anechoic chamber at similar physical temperature as the instrument. The Corbella equation has been verified from ground, airborne and space.

The flat target transformation

According to the Corbella equation, any uncertainty in the knowledge of the antenna pattern is amplified by the instrument–target temperature contrast. Thus, it is desirable to minimize such contrast. This is possible if the *flat target response* (FTR) of the interferometer is acquired

first. The FTR is the visibility function measured when imaging an unpolarized uniform brightness temperature distribution, which is stable over time. In practice, the cold sky near the galactic poles of the Milky Way is a good approximation of such a target, and the FTR can be measured by pointing the aperture synthesis radiometer to the galactic poles. The FTR is like a mask formed by ripples due to system imperfections, mainly antenna errors. When this mask is scaled down to the temperature contrast between the instrument and an Earth scene, and is subtracted from the measured visibilities, the instrument errors are cancelled to a large extent. Mathematically, the Corbella equation is transformed by this linear combination with the result that the instrument physical temperature is replaced by the average temperature of the scene. The final temperature contrast is then reduced to that between each pixel of the scene and the average of all the pixels. The transformation of the Corbella equation using the FTR is known as the *flat target transformation* (FTT) and is an essential step in aperture synthesis imaging (Martín-Neira et al., 2008).

Polarimetry in aperture synthesis

An aperture synthesis radiometer is inherently a polarimetric type of instrument because of its very wide field of view. Horizontal H and vertical V polarizations on the target are transformed into v and h polarizations on the element antenna according to a different rotation matrix depending on the direction. Since the H and V fields appear mixed together in the v and h fields in the instrument polarization frame, a nonzero vh cross brightness temperature is measured even from a target with no correlation between H and V . Reversely, to recover the horizontal and vertical brightness temperatures of the target, it is necessary to measure the vh cross brightness temperatures. When these are not available, a singularity along the image diagonals appears, and pixels in those areas have to be discarded. They can be recovered though when, due to the platform motion, they move to directions outside the singularities.

Calibration

An aperture synthesis radiometer is a complex instrument and requires several calibration steps. The first one is the flat target transformation, explained above, for which a cold sky view to acquire the flat target response must have been performed before hand. The second step is the correction of any comparator offset in the digital circuits and quadrature error of the demodulator. Then comes the phase correction, achieved by injecting correlated noise to all receivers internally, and the amplitude calibration, based on two standard temperatures, as the cold sky and an internal matched load at a well-monitored physical temperature. Finally, an uncorrelated load is used to remove any correlation offsets due to residual internal interference. In addition, to form an image, the antenna patterns must have been characterized a priori, and the

fringe-washing function has to be determined during operation. The fringe-washing function takes into account the decorrelation of the signal with time delay across the array and is estimated from early, punctual, and late correlations.

Summary and conclusions

Aperture synthesis radiometry has developed starting in the early 1980s mainly driven by the need to map soil moisture and ocean salinity from space. During this development, the theoretical fundamentals had to be revised, and a new formulation, the Corbella equation, was found, which describes correctly the operation of this type of instruments, and, in fact, of radio telescopes as well. A consequence of the new formulation is the flat target transformation, a powerful method to calibrate out antenna errors in aperture synthesis.

The spatial resolution and sensitivity of an aperture synthesis radiometer are comparable to those of a real aperture radiometer. Field of view, collecting area, and integration time take different values in each case, but balance each other, yielding the same net result at the end.

ESA's SMOS mission carrying the first aperture synthesis radiometer into space was launched 2 November 2009. Scientists have been able to assess the ultimate benefits of this new remote sensing technique. Following SMOS success, a next generation of interferometric radiometers might follow not only from low Earth orbit but also from geostationary orbit, already under study.

Bibliography

- Anterrieu, E., Waldteufel, P., and Lannes, A., 2002. Apodization functions for 2-D hexagonally sampled synthetic aperture imaging radiometers. *IEEE Transaction in Geoscience and Remote Sensing*, **40**(12), 2531–2542.
- Bayle, F., Wigneron, J.-P., Kerr, Y. H., Waldteufel, P., Anterrieu, E., Orllhac, J.-C., Chanzy, A., Marloie, O., Bernardini, M., Sobjaerg, S., Calvet, J.-C., Goutoule, J.-M., and Skou, N., 2002. Two-dimensional synthetic aperture images over a land surface scene. *IEEE Transaction in Geoscience and Remote Sensing*, **40**(3), 710–714.
- Camps, A., 1996. *Application of Interferometric Radiometry to Earth Observation*. PhD thesis, Barcelona, Polytechnic University of Catalonia.
- Christensen, J., Carlstrom, A., Ekstrom, H., Emrich, A., Embretsen, J., De Maagt, P., and Colliander, A., 2007. GAS: the geostationary atmospheric sounder. In *IGARSS-2007 Proceedings*, Barcelona, pp. 223–226.
- Colliander, A., Lemmetyinen, J., Uusitalo, J., Suomela, J., Veijola, K., Kontu, A., Kemppainen, S., Pihlflyckt, J., Rautiainen, K., Hallikainen, M., and Lahtinen, J., 2007. Ground calibration of SMOS: NIR and CAS. In *IGARSS Proceedings*, pp. 3631–3634.
- Corbella, I., Duffo, N., Vall-llossera, M., Camps, A., and Torres, F., 2004. The visibility function in interferometric aperture synthesis radiometry. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(8), 1677–1682.
- Goutoule, J. M., Kraft, U., and Martín-Neira, M., 1994. MIRAS: preliminary concept of a two-dimensional L-band aperture synthesis radiometer. In *MicroRad'94 Proceedings*, Rome.
- Kainulainen, J., Rautiainen, K., Tauriainen, S., Auer, T., Kettunen, J., and Hallikainen, M., 2007. First 2D interferometric

- radiometer imaging of the Earth from an aircraft. *IEEE Transactions on Geoscience and Remote Sensing Letters*, **4**(2), 241–245.
- Kerr, Y. H., Wigneron, J. P., Ferrazzoli, P., and Waldteufel, P., 2000. Soil moisture and vegetation biomass retrievals using L-band, dual polarised and multi angular radiometric data in preparation of the SMOS mission. In *IGARSS-2000 Proceedings*, Hawaii, Vol. 3, pp. 1244–1246.
- Komiyama, K., 1990. *Super-synthesis radiometer (SSR) for the remote sensing of the Earth*. Technical Report. Ibaraki: Electrotechnical Laboratory.
- Krauss, J. D., 1996. *Radio Astronomy*. New York: McGraw-Hill.
- Lannes, A., and Anierrieu, E., 1994. Image reconstruction methods for remote sensing by aperture synthesis. In *IGARSS-94 Proceedings*, Pasadena, Vol. 4, pp. 2228–2230.
- Laursen, B., and Skou, N., 1994. A spaceborne synthetic aperture radiometer simulated by the TUD demonstrator model. In *IGARSS-94 Proceedings*, Pasadena, Vol. 3, pp. 1314–1316.
- LeVine, D. M., and Good, J. C., 1983. *Aperture synthesis for microwave radiometers in space*. NASA Technical Memorandum 85033. Greenbelt, MD: Goddard Space Flight Center.
- LeVine, D. M., Griffiths, A., Swift, C. T., and Jackson, T. J., 1992. ESTAR: a synthetic aperture microwave radiometer for measuring soil moisture. In *IGARSS-92 Proceedings*, Houston, TX, Vol. II, pp. 1755–1757.
- Martín-Neira, M., 1993. MIRAS: a two-dimensional passive aperture synthesis radiometer. In *Presentation at PIERS'93*. Pasadena, CA: JPL.
- Martín-Neira, M., Suess, M., Kainulainen, J., and Martín-Porqueras, F., 2008. The flat target transformation. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(3), 613–620.
- McMullan, K., Brown, M., Martín-Neira, M., Rits, W., Ekholm, S., Marti, J., and Lemanczyk, J. 2008. SMOS: the payload. *IEEE TGARS*, **46**, 594–605.
- Rautiainen, K., Valmu, H., Jukkala, P., Moren, G., and Hallikainen, M., 1999. Four-element prototype of the HUT interferometric radiometer. In *IGARSS-99 Proceedings*, Hamburg, Vol. 1, pp. 234–236.
- Ruf, C. S., 1990. Antenna performance for a synthetic aperture microwave radiometer in geosynchronous Earth orbit. In *IGARSS-90 Proceedings*, pp. 1589–1592.
- Ruf, C. S., Swift, C. T., Tanner, A. B., and LeVine, D. M., 1988. Interferometric synthetic aperture microwave radiometry for the remote sensing of the Earth. *IEEE Transactions on Geoscience and Remote Sensing*, **26**(5), 597–611.
- Ryle, M., Smith, F. G., and Elsmore, B., 1950. *A Preliminary Survey of the Radio Stars in the Northern Hemisphere*. London: Royal Astronomical Society.
- Schanda, E., 1979. Multiple wavelength aperture synthesis for passive sensing of the Earth's surface. In *International Symposium Digest*. Seattle, WA: IEEE Antennas and Propagation Society, p. 762.
- Skou, N., 2004. Spaceborne L-band radiometers: pushbroom or synthetic aperture? In *IGARSS-04 Proceedings*, Vol. 2, pp. 1264–1267.
- Swift, C. T., 1993. *ESTAR – The Electronically Scanned Thinned Array Radiometer for Remote Sensing Measurement of Soil Moisture and Ocean Salinity*. Goddard Space Flight Center, Washington DC: NASA, p. 4523.
- Swift, C. T., LeVine, D. M., and Ruf, C. S., 1991. Aperture synthesis concept in microwave remote sensing of the Earth. *IEEE Transactions on Microwave Theory and Techniques*, **39**, 1931–1935.
- Tanner, A. B., 1990. *Aperture Synthesis for Passive Microwave Remote Sensing: The Electronically Scanned Thinned Array Radiometer*. PhD thesis, Amherst, University of Massachusetts.
- Thompson, A. R., Moran, J. M., and Swenson, G. W., 1988. *Interferometry and Synthesis in Radio Astronomy*. New York: Wiley.
- Torre, F., Camps, A., Bara, J., Corbella, I., and Ferrero, R., 1996. On-board phase and modulus calibration of large aperture synthesis radiometers: study applied to MIRAS. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(4), 1000–1009.
- Wedge, S. W., and Rutledge, D. B., 1991. Noise waves and passive linear multiports. *IEEE Microwave and Guided Wave Letters*, **1**(5), 117–119.
- Wu, J., Liu, H., Yan, J., Ban, S., Dong, X., and Jiang, J., 2005. Research activity on synthetic aperture radiometry in CSSAR/CAS. In *PIERS 2005 Proceedings*, Hangzhou.
- Yan, J., Wu, J., Liu, H., Dong, X., and Jiang, J., 2005. Design and implementation of digital correlator for CAS synthetic aperture radiometer. In *PIERS 2005 Proceedings*, Hangzhou.

Cross-references

[Microwave Radiometers](#)
[Microwave Radiometers, Conventional](#)
[Microwave Radiometers, Correlation](#)
[Microwave Radiometers, Polarimetry](#)
[Observational Systems, Satellite](#)
[Sea Surface Salinity](#)
[Soil Moisture](#)

MICROWAVE RADIOMETERS, POLARIMETERS

David Kunkee
 The Aerospace Corporation, Los Angeles, CA, USA

Synonyms

Microwave polarimeter; Polarimetric radiometer

Definition

Microwave Polarimetric Radiometer. Microwave radiometer capable of measuring at least three of the four modified Stokes' parameters of the incident field.
Polarimetry. Measurement of the polarization characteristics of incident electromagnetic radiation.

Introduction

Polarimetric microwave radiometers are capable of measuring at least three Stokes' parameters of the incident microwave radiation or radiometric scene. Traditional dual-polarized radiometers measure vertically and horizontally polarized brightness temperatures or, equivalently, the first two Stokes' parameters of the input scene. In order to completely characterize the spatially averaged input scene, measurements of third and fourth Stokes' parameters are required (Jackson, 1975), and these measurements can be performed utilizing coherent (direct) or incoherent techniques. Direct measurements are performed by polarization-correlating radiometers to measure the third and/or fourth Stokes' parameters. Indirect methods include a combination of linearly polarized brightness temperatures typically oriented at 45° and –45° with respect to the principle directions in order to derive the third Stokes' parameter and by combining left-and right-hand circular polarized brightness temperatures to derive the fourth

Stokes' parameter. Several airborne microwave polarimeters have been part of airborne experiments conducted between 1993 and 2003 in order to investigate application of microwave polarimetry to retrieval of sea surface wind direction (Yueh et al., 1995). The satellite-based wind direction system (WindSat), developed by the US Naval Research Laboratory (NRL) and launched in January 2003, performed the first space-based polarimetric measurements (Gaiser et al., 2004).

Background

Airborne and space-based polarimetric radiometers measure the intensity of the Earth's naturally occurring upwelling thermal emission. This arbitrarily polarized, non-or partially coherent thermal radiation is completely characterized by the four parameter-modified Stokes' vector:

$$\bar{I} = \begin{pmatrix} I_v \\ I_h \\ U \\ V \end{pmatrix} \frac{W}{(m^2 \cdot sr \cdot Hz)}$$

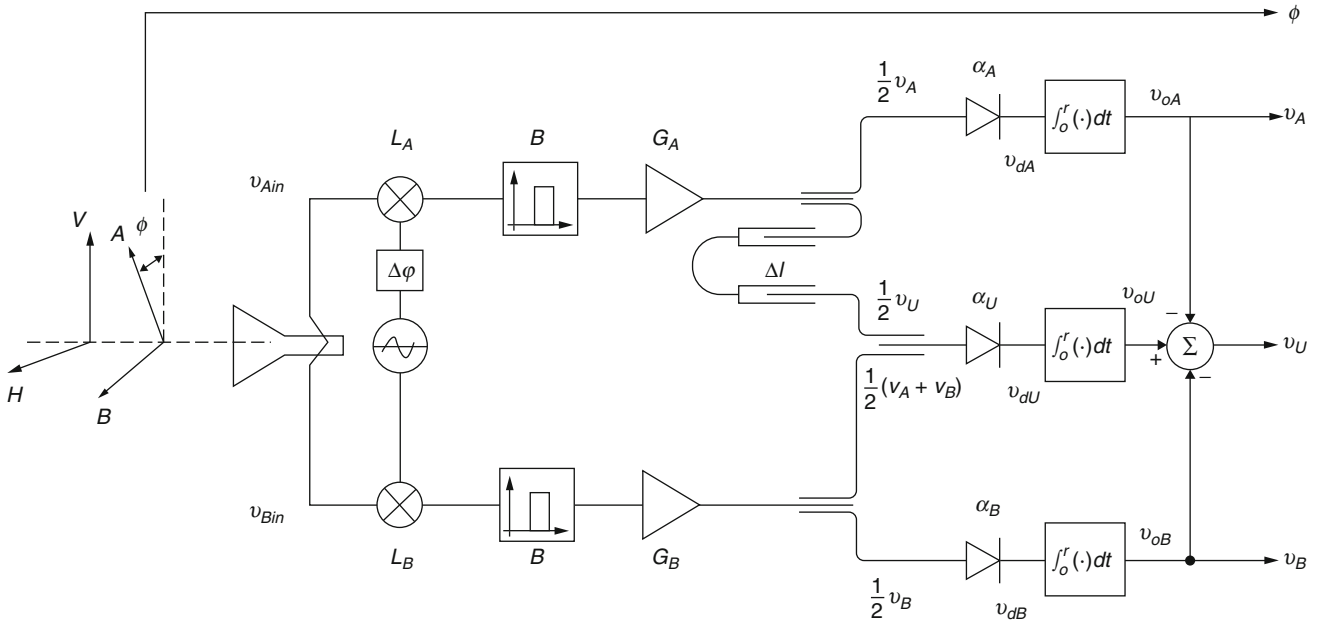
where I_v and I_h are the spectral intensities of the vertically and horizontally polarized field components and U and V are the in-phase and quadrature correlations, respectively, between the orthogonally polarized field components $\hat{v}$ and $\hat{h}$. The spectral intensity of this emission can be described by the brightness temperature vector, $\bar{T}_B$:

$$\bar{T}_B \equiv \frac{\lambda^2}{k} \bar{I} = \begin{pmatrix} T_v \\ T_h \\ T_U \\ T_V \end{pmatrix} = \begin{pmatrix} \langle |E_v|^2 \rangle \\ \langle |E_h|^2 \rangle \\ 2\text{Re}\{\langle E_v E_h^* \rangle\} \\ 2\text{Im}\{\langle E_v E_h^* \rangle\} \end{pmatrix}$$

where E_v and E_h represent the vertically and horizontally polarized electric fields, respectively, $\langle \cdot \rangle$ represents a time-averaged quantity. Microwave polarimeters measure at least three of the four parameters. There are two general methods in passive microwave remote sensing in order to measure the third and fourth Stokes' parameter: (1) cross-correlation between two orthogonally polarized measurements (polarization-correlating radiometer) and (2) measurements of polarized brightness temperature measurements using multiple polarizations. For example, using linearly polarized brightness temperatures at orientations of $\pm 45^\circ$ with respect to the orientation of the vertical polarization vector, the third Stokes' parameter can be found simply by differencing the two measurements: $T_U = T_{45} - T_{-45}$. Similar approaches can be used to measure T_V . One of the first reports on radiometer designs for sea surface microwave emission polarimetry was provided by Dzura (1992).

Polarization-correlating radiometer (direct method)

An example of an analog polarization-correlating radiometer is given below. Figure 1 shows the block diagram of a three-channel polarization-correlating radiometer.



Microwave Radiometers, Polarimeters, Figure 1 Block diagram of a three-channel polarization-correlating radiometer (Gasiowski and Kunkee, 1993).

The measurement sensitivity for v_U (and therefore T_U) is proportional to the geometric mean of the characteristics of v_A and v_B ; hence, if channels A and B are identical, channel U will have the same sensitivity (ΔT_{RMS}) as channels A and B. Phase adjustments are made by adjusting the relative oscillator phase ($\Delta\phi$) and/or relative line length (Δl) before the combiner. The phase can be adjusted to provide response to in-phase (U) or phase quadrature (V) components; however, in order to measure V (fourth Stokes' parameter), a single-side band configuration must be used in order to avoid cancelation of the signal from the contributions of the high and low side bands.

The post-detection analog signal summation in the scheme shown in Figure 1 is typically performed by using digital electronic circuitry (Piepmeier and Gasiewski, 2001).

Polarization combining radiometer (indirect method)

An alternate method is to combine brightness temperature measurements at multiple polarizations to derive U or V indirectly. This method is also called the incoherent method and is typically applied through measurements of linearly polarized brightness temperatures offset $\pm 45^\circ$ from the original basis to derive the third Stokes' parameter T_U and using left- and right-hand-polarized brightness temperatures to derive the fourth Stokes' parameter, T_V . Using a unitary rotational transform, the relationship between linearly polarized brightness temperatures rotated ϕ degrees is expressed by

$$\begin{aligned} \overline{T}' &= \begin{pmatrix} T_{\nu'} \\ T_{h'} \\ T_{U'} \\ T_{V'} \end{pmatrix} \\ &= \begin{bmatrix} \cos^2 \phi & \sin^2 \phi & 0.5 \sin^2 \phi & 0 \\ \sin^2 \phi & \cos^2 \phi & -0.5 \sin^2 \phi & 0 \\ -\sin^2 2\phi & \sin^2 2\phi & \cos 2\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} T_{\nu} \\ T_h \\ T_U \\ T_V \end{pmatrix} \\ &= \overline{U}(\phi) \overline{T} \end{aligned}$$

where $\overline{T}'$ is the measurement made with the rotated polarization basis. Therefore, by using $T_{\pm 45} = T_{U'}(\phi = \pm 45^\circ)$, the third Stokes' parameter, $T_U = T_{45} - T_{-45}$. A popular way of deriving brightness temperature measurements at $\pm 45^\circ$ is through the use of polarization combining networks (Yueh et al., 1995). Note that in the above equation, the fourth Stokes' parameter, T_V , is rotationally invariant; however, it can be shown that $T_V = T_R - T_L$ where T_R and T_L are the right- and left-hand circularly polarized brightness temperatures of the incident wave, respectively (Tsang et al., 1985).

Ocean surface wind measurements

One of the primary applications of microwave polarimetry is to improve ocean surface wind (OSW) retrievals

performed with dual-polarized, T_{ν} and T_h , brightness temperatures at 18 and 37 GHz. Dual-polarized radiometric measurements from space-based radiometers are currently and routinely used to accurately observe ocean surface wind speed to better than 1 m/s accuracy, Wentz (1997). The relationship between wind-roughened ocean surfaces and upwelling brightness temperature is well understood and was first explored by Hollinger (1971). Ocean surface wind speed measurements have been performed from operational microwave radiometers since the first special sensor microwave/imager (SSM/I) radiometer was launched and began operating in 1987 (Goodberlet et al., 1990; see also "*Ocean, Measurements and Applications*").

The first measurements showing a variation of brightness temperature as a function of the relative angle between the surface wind direction and the polarization plane were reported by Bepalova et al. (1979) as part of an airborne radiometry experiment conducted by the Space Research Institute at Moscow. A study by Wentz (1992) showed that there was a systematic bias in wind speed retrievals from SSM/I compared to ocean buoys that was dependent on the angle of observation with respect to the surface wind direction. The directional dependence was used to demonstrate that wind direction could also be retrieved using microwave brightness temperatures. However, the presence of multiple solutions, or directional ambiguities, limits the utility of wind direction derived using dual-polarized microwave brightness temperatures.

Building upon theoretical modeling and laboratory measurements of polarized microwave emission from wave-covered surfaces (Yueh et al., 1994; Gasiewski and Kunkee, 1994), the utility of the third Stokes' parameter for ocean surface wind direction retrieval was demonstrated through airborne experiments that measured polarimetric brightness temperatures over the ocean (Yueh et al., 1995). In order to demonstrate wind direction retrievals from space, the WindSat microwave polarimetric radiometer was developed by the US Naval Research Laboratory and launched in 2003 (Gaiser et al., 2004). Wind vector (speed and direction) retrievals using WindSat data (Bettenhausen et al., 2006) are now assimilated into numerical weather forecasting (NWP) models and used to improve forecasts (Candy et al., 2009).

Other applications

Measurements of the third and fourth Stokes' parameters by WindSat over the Greenland ice sheet have shown responses related to the asymmetrical features of polar ice sheets and potentially the microphysical structure of snow crystals (Li et al., 2008). Over desert, polarimetric measurements are also sensitive to the asymmetrical features of the surface such as the structure of sand dunes (Narvekar et al., 2007). Investigations have also shown the WindSat polarimetric measurements to be highly sensitive to RFI signals, and in some cases, showing responses before contamination is apparent in the vertically or horizontally polarized channels (Ellingson and Johnson, 2006; see also "*RFT*").

Summary

Microwave polarimetry involves the measurement of the third and fourth Stokes' parameter of the input radiometric "scene." Traditional dual-polarized radiometers measure only the first two parameters of the Stokes' vector and miss some characteristics of the brightness temperature scene. Microwave polarimeters can be designed for direct (coherent) polarimetric measurements or indirect (incoherent) measurements that are derived through post-detection combination of polarized measurements. Motivation to develop airborne and space-based microwave polarimeters is provided by improvements to passive ocean surface wind vector measurements. The WindSat radiometer was the first space-based microwave polarimeter, providing global measurements of all four Stokes' parameters. These measurements are used to retrieve ocean surface wind vectors and improve weather forecast quality of NWP models. Other potential applications of microwave polarimetry are under development and include snow and ice characterization, land monitoring, and RFI detection and characterization.

Bibliography

- Bespalova, E. A., Veselov, V. M., Glotov, A. A., et al., 1979. Sea ripple anisotropy estimates from variations in polarized thermal emission of the sea. *Doklady Akademii Nauk SSSR*, **246**(6), 1482–1485.
- Bettenhausen, M. H., Smith, C. K., Bevilacqua, R. M., Wang, N., and Gaiser, P. W., 2006. A nonlinear optimization algorithm for WindSat wind vector retrievals. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(3), 597–609.
- Candy, B., English, S. J., and Keogh, S., 2009. A comparison of the impact of quickscat and windsat wind vector products on met office analyses and forecasts. *IEEE Transactions on Geoscience and Remote Sensing*, **47**(6), 1632–1640.
- Dzura, M. S., Etkin, V. S., Khrupin, A. S., Pospelov M. N., and Raev M. D., 1995. Radiometers-polarimeters: principles of design and applications for sea surface microwave emission anisotropy. In *IEEE International Geoscience and Remote Sensing Symposium*, 1995. IGARSS'95, Vol. 2, pp. 1432–1434.
- Ellingson, S. W., and Johnson, J. T., 2006. A polarimetric survey of radio frequency interference in C-and X-bands in the continental United States using windsat radiometry. *IEEE Transaction on Microwave Theory and Techniques*, **44**(3), 540–548.
- Gaiser, P. W., et al., 2004. WindSat spaceborne polarimetric radiometer: sensor description and early orbit performance. *IEEE Transaction on Microwave Theory and Techniques*, **42**(11), 2347–2361.
- Gasiewski, A. J., and Kunkee, D. B., 1993. Calibration and applications of polarization correlating radiometers. *IEEE Transaction on Microwave Theory and Techniques*, **41**(5), 767–773.
- Gasiewski, A. J., and Kunkee, D. B., 1994. Polarized microwave emission from water waves. *Radio Science*, **29**(6), 1449–1466.
- Goodberlet, M. A., Swift, C. T., and Wilkerson, J. C., 1990. Ocean surface wind speed measurements of the special sensor microwave/imager (SSM/I). *IEEE Transactions on Geoscience and Remote Sensing*, **28**(5), 823–828.
- Hollinger, J. P., 1971. Passive microwave measurements of sea surface roughness. *IEEE Transactions on Geoscience Electronics*, **GE-9**(3), 165–169.
- Jackson, J. D., 1975. *Classical Electrodynamics*. New York: Wiley, p. 848.
- Li, L., Gaiser, P. W., Albert, M. R., Long, D. G., and Twarog, E. M., 2008. WindSat passive polarimetric signatures of the greenland ice sheet. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(9), 2622–2631.
- Narvekar, P. S., Jackson, T. J., Bindlish, R., Li, L., Heygster, G., and Gaiser, P. W., 2007. Observations of land surface passive polarimetry with the windsat instrument. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(7), 2019–2028.
- Piepmeyer, J. R., and Gasiewski, A. J., 2001. Digital correlation microwave polarimetry: analysis and demonstration. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(11), 2392–2410.
- Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley, p. 613.
- Wentz, F. J., 1992. Measurement of oceanic wind vector using satellite microwave radiometers. *IEEE Transactions on Geoscience and Remote Sensing*, **30**(5), 960–972.
- Wentz, F. J., 1997. A well-calibrated ocean algorithm for special sensor microwave/imager. *Journal of Geophysical Research*, **102**(C4), 8703–8718.
- Yueh, S. H., Nghiem, S. V., Kwok, R., Wilson, W. J., Li, F. K., Johnson, J. T., and Kong, J. A., 1994. Polarimetric thermal emission from periodic water surfaces. *Radio Science*, **29**, 87–96.
- Yueh, S. H., Wilson, W. J., Li, F. K., Nghiem, S. V., and Ricketts, W. B., 1995. Polarimetric measurements of sea surface brightness temperatures using an aircraft K-band radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, **33**(1), 85–92.

MICROWAVE SUBSURFACE PROPAGATION AND SCATTERING

Alexander Yarovoy

Delft University of Technology, Delft, The Netherlands

Definitions

Subsurface. Natural materials (soils, rock, snow, ice) below air–ground interface.

Ground-penetrating radar. Radar for subsurface sensing. It locates, images, and characterizes changes in electrical and magnetic properties of subsurface materials.

Introduction

The first description of microwaves use for subsurface sensing is attributed to a German patent by Leimbach and Löwy from 1910 (Daniels, 2004). In this patent, propagation of microwaves between pairs of vertically buried dipole antennas has been used to detect any subsurface objects with higher conductivity than the surrounding medium. Only monochromatic electromagnetic waves have been considered in this patent. The first use of electromagnetic pulses with a broad spectrum to determine the structure of buried objects is attributed to Hülsenbeck (Hülsenbeck et al., 1926). It was noted that any dielectric variation, not necessarily involving conductivity, would also produce reflections. The first ever experiments with subsurface microwave propagation to sound the depth of a glacier were performed in Austria in 1929 (Stern, 1929, 1930). These experiments were largely forgotten

until the late 1950s when US Air Force radars were seeing through ice as planes tried to land in Greenland but misread the altitude and crashed into the ice. This accident together with lunar soil investigations within Apollo program (Simmons et al., 1972) triggered research on microwave penetration into subsurface, including not only ice sounding but also mapping soil properties and the water table. This fundamental research and simultaneous development of ultra-wideband microwave technology resulted, in the early 1970s, in wide use of microwave subsurface propagation and scattering for characterization and mapping of subsurface. Since then, the range of applications has been expanding steadily.

Dielectric properties of soil, rock, ice, and snow

Propagation velocity and attenuation of microwaves is controlled by the electrical and magnetic properties of subsurface. Regarding the subsurface, three different types of natural earth materials should be distinguished: rocks, soils, and ice/packed snow. At microwave frequencies, electrical properties (which in most cases are more important than the magnetic properties) are dominantly controlled by density and by the chemistry, fine structure (composition of liquid/gas/solid components), and content of water. Below we concentrate mainly on rock and soil properties.

Following Heimovaara et al. (1994), soils can be considered as a four-component system: irregularly shaped solid particles, air, free water, and bound water. The bound water refers to the first few (up to 10) molecular layers of water near solid surfaces that are rotationally hindered by surface forces. The frequency-dependent complex dielectric permittivity of soils can be described then with a four-component complex dielectric mixing model based on the volumetric mixing of the refractive indices of the soil components. It should be noted that free water and bound water have different dielectric permittivities and thus different impact of the electrical properties of the material. Similarly to this approach, Wobschall (1977) has considered rocks as a three-phase system: irregularly shaped particles, air-or water-filled voids (pores), and crevices. In both cases, presence of moisture strongly influences electrical properties of material.

Solid particles of soils and rocks have different physical properties depending on their material. In the most simple and frequently used approach (Wang and Schmugge, 1980; Dobson, et al., 1985), the soil solid particles are considered to be a mixture of sand particles of diameter $d > 0.005$ cm; silt, 0.0002 cm $< d < 0.005$ cm; and clay, $d < 0.0002$ cm, weight content of which is expressed in percentage of total weight of soil. Dielectric permittivities of dry sand, clay, and silt are slightly different (see Table 1). More important is that clay, silt, and sand have different abilities to bind the water. It is shown in Wang and Schmugge (1980) that the quantity of bound water in soil depends on the volume of clay fraction in it and

Microwave Subsurface Propagation and Scattering,
Table 1 Typical values of relative dielectric permittivity of some dry soils and rocks at microwave frequencies (Daniels, 2004)

Material	Relative dielectric permittivity
Sand	2.4–6
Loam	4–10
Clay	2–6
Granite	5
Limestone	7
Sandstone	2–5
Silt	3–10

the quantity of bound water increases with the volume of clay. This is explained by a large specific area of clay surface compared to other soil fractions. Sand and silt particles are also covered with bound water films; however, the amount of bound water on sand and silt particles is less than 0.1 % (Bojarskii et al., 2002). Further discussions on dielectric properties of soils and rocks can be found in Hoekstra and Delaney (1974), Hipp (1974), De Loor (1983), and Hallikainen et al. (1985).

Providing a deep insight in the interaction of microwaves with rocks and soils, the abovementioned approaches require detailed knowledge of many physical parameters of soils, which are often not known during the field work. Thus, a large number of approximate equations have been proposed to compute dielectric permittivity of soil as a mixture of different components (Shutko and Reutov, 1982; Dobson et al., 1985; Sen et al., 1981). Among them, the so-called Bruggeman–Hanai–Sen (BHS) mixing model (Sen et al., 1981) is widely used nowadays:

$$p = \frac{(\epsilon_m - \epsilon_r) \left(\frac{\epsilon_w}{\epsilon_r} \right)^C}{\epsilon_m - \epsilon_w}$$

where ϵ_r is bulk soil relative dielectric permittivity, ϵ_m is the dielectric permittivity of dry soil, ϵ_w is the dielectric permittivity of water, p is fractional porosity (volume of voids/total volume), and C is a shape factor (1/3 for spherical grains). When the water content is large, wet rock or soil can be considered simply as dry solid/water mixture and simply empirically stated Topp equation (Topp et al., 1980) is frequently used:

$$\epsilon_r = 3.03 + 9.3 \cdot \theta + 146.0 \cdot \theta^2 - 76.3 \cdot \theta^3$$

where ϵ_r is bulk soil relative dielectric permittivity and θ is volumetric water content. Increase of the water content in general is responsible for increase of relative dielectric permittivity of the material.

In order to use different dielectric mixture formulas, knowledge of dielectric permittivity of water and dry soils is required. The complex-valued dielectric permittivity of bulk water at microwave frequencies can be described by

the Debye theory and the Cole–Cole model (Cole and Cole, 1941; Heimovaara, 1994) is frequently used:

$$\varepsilon(\omega) = \varepsilon' - j\varepsilon'' = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_\infty}{1 + j(\omega\tau)^{1-\beta}} - j\frac{\sigma_{dc}}{\omega\varepsilon_0}$$

where ε_s is the low-frequency (static) dielectric permittivity of water (which is of about 88 at a temperature of 0 °C), ε_∞ is the high (infinite)-frequency dielectric permittivity and equals 4.25 for the 0 °C, τ is the relaxation time of water and $\beta = 0.0125$ and is a factor that accounts for the possible spread in relaxation frequencies, σ_{dc} is the DC conductivity at a given temperature, and ε_0 is the permittivity of free space, which is $8.854 \cdot 10^{-12}$ F/m. The relaxation time of water molecules, which equals $7.7 \cdot 10^{-12}$ s at 27 °C, drastically increases due to surface forces, and in monomolecular layers of water covering, the soil particles can reach the value of $5.0 \cdot 10^{-10}$ s at 27 °C (Bojarskii et al., 2002). As a result, the bound water exhibits relaxation at frequencies below 300 MHz and has, at microwave frequencies, considerably lower values for the dielectric permittivity than free water. The permittivity of bound water is temperature dependent, resulting in a substantial increase in bulk dielectric permittivity with temperature at microwave frequencies for soils with high surface area (Or and Wraith, 1999).

In an idealized representation, the water relative permittivity remains constant at high (roughly above 10 GHz) and low (below 100 MHz) frequencies. In a transition region over a frequency band, the real part of complex dielectric permittivity considerably decreases, while imaginary part exhibits a maximum.

Solid parts of most soils and rocks have, when dry, a relative permittivity in the range 2–9 (see Table 1) (Daniels, 2004). Very often, it is assumed that the relative permittivity of dry soils and rock is frequency independent at microwave frequencies.

Total losses in natural rocks and soils are mainly due to conductivity and polarization losses. According to Olhoeft (see, e.g., Olhoeft, 1987), water plays the dominant role in rock and soil conductivity through ionic charge transport through water-filled pore spaces in rocks and soils. Furthermore, conductivity might be due to presence of metals and other good conductors. Conductivity results in dissipation of microwave energy into heat.

Similarly to soils, dielectric properties of snow and seawater ice depend on texture and volumetric amount of water and ice crystal (Bogorodskii et al., 1983; Boyarskii and Tikhonov, 1994; Kovacs et al., 1995). Relaxation frequencies of water molecule in ice crystals lie in kHz region. Thus at microwave frequencies, pure ice has almost constant with frequency dielectric permittivity of about 3.17 and very low losses. Dry snow can be considered as a mixture of ice crystals and air. Depending on density of snow, its relative dielectric permittivity varies from 1.2 to 2. Dielectric permittivity of wet snow depends largely on amount of water in it.

Seawater ice has different properties than freshwater ice, because seawater ice consists of pure ice, air, brine, and possibly solid salts. Furthermore, the sea ice has a multilayer structure, which depends on periodic processes of growth, deformation, and melting. Depending on its age and internal structure, several types of sea ice are distinguished. Relative dielectric permittivity of ice may vary from 3 till 18–30. Losses of coherent energy of microwaves in ice are determined by ice conductivity and noncoherent scattering on spatial heterogeneities.

Attenuation and dispersion of microwaves in subsurface

To simplify the mathematical description of microwave propagation in natural subsurface, a complex-valued apparent dielectric permittivity is used:

$$\varepsilon = \varepsilon' - j\varepsilon''$$

where ε' is a bulk dielectric permittivity of the subsurface and ε'' stands for losses associated with both conductivity and polarization (dipolar) losses:

$$\varepsilon'' = \frac{\sigma}{j\omega} + \varepsilon''_{relaxation}$$

The polarization losses are caused mainly by dielectric relaxation of water and are due to the polarity of the two hydrogen atoms of the water molecule.

The plane electromagnetic wave propagating in such medium can be presented as a function of distance z and time t :

$$E(z, t) = E_0 \cdot e^{-\alpha z} \cdot e^{j(\omega t - \beta z)}$$

where E stands for the electric field component of the field, α is the attenuation factor, and β is the phase constant. Both parameters are determined by the complex-valued apparent dielectric permittivity of the medium:

$$\alpha = \omega \left[\frac{\mu\varepsilon'}{2} \left(\sqrt{1 + \left(\frac{\varepsilon''}{\varepsilon'} \right)^2} - 1 \right) \right]^{\frac{1}{2}}$$

$$\beta = \omega \left[\frac{\mu\varepsilon'}{2} \left(\sqrt{1 + \left(\frac{\varepsilon''}{\varepsilon'} \right)^2} + 1 \right) \right]^{\frac{1}{2}}$$

The distance d at which the electric field is attenuated in e times is called the skin depth and is equaled to $1/\alpha$. The skin depth is widely used in practice to estimate penetration depth of microwaves into subsurface or thickness of the top subsurface layer which determines microwave emission from the subsurface. It can be seen that the attenuation factor α by ignoring frequency dependence of complex-valued apparent dielectric permittivity is linearly related to frequency. Thus, the skin depth in the very first approximation also linearly decreases with frequency.

Attenuation of microwaves in dry subsurface at 100 MHz varies from typically less than 1 dB/m for snow and sand up to 10 dB/m for some clay, limestone, and sandstone (Daniels, 2004). The attenuation factor slowly increases with frequency in the range from 100 MHz till 1 GHz, while at the frequencies above 1 GHz, it sharply increases due to dielectric relaxation of water. At, for example, 3 GHz, microwave attenuation in subsurface varies typically between several dozens and several hundreds dB/m.

The phase constant β determines the phase velocity of microwave propagation in subsurface:

$$v = \frac{\omega}{\beta} = \left[\frac{\mu\epsilon'}{2} \left(\sqrt{1 + \left(\frac{\epsilon''}{\epsilon'} \right)^2} + 1 \right) \right]^{-\frac{1}{2}}$$

In the case of small losses and far away from relaxation frequencies, the microwave propagates with the phase velocity:

$$u = \frac{c}{\sqrt{\epsilon_r}} = \frac{1}{\sqrt{\mu_0\epsilon_0\epsilon_r}}$$

which is frequency independent. In the general case, finite conductivity of subsurface and the dielectric relaxation of water are responsible for dependence of phase velocity on frequency. This dependence is called dispersion. For microwaves, the dispersion is typically observed at the frequencies above 1 GHz.

By propagation of wideband microwave pulses in subsurface, the dispersion and attenuation cause not only decrease of the pulse magnitude but also distortion of the pulse waveform.

Subsurface scattering

Subsurface is heterogeneous not only on microscopic but also on macroscopic scale. On top of regular geological structure created by multiple strata of different soils/rocks, there is a finer irregular structure created by local spatial inhomogeneities. These are created by spatial variations of moisture, all kind of inclusions (e.g., stones or man-made objects), and animal burrows. While propagating in subsurface, microwaves scatter on interfaces between different strata as well as on local inhomogeneities. Fields scattered from large-scale heterogeneities are used in subsurface remote sensing for geological structure reconstruction and minerals prospecting (Fung, 1994; Daniels, 2004). It has been shown by Yarovoy et al. (2000) that the presence of a low-loss (e.g., sand) layer above rock at certain angles and polarizations of the incident field will enhance microwave backscattering from either internal interface (rock surface) or from air-ground interface.

Field scattered on small-scale (local) inhomogeneities is called subsurface clutter. At medium and/or high microwave frequencies, typical size of many abovementioned local inhomogeneities becomes

comparable with the wavelength of electromagnetic field. As a consequence, strength of clutter at these frequencies might become comparable to useful (searched) reflections, resulting in masking these reflections. It is widely believed that at frequencies above 1 GHz, the subsurface clutter becomes the major limiting factor for subsurface sensing.

Subsurface sensing

Subsurface propagation and scattering of microwaves makes a basis for subsurface sensing. Nowadays, three major approaches are used: ground-penetrating radar (GPR) and airborne and spaceborne remote sensing.

GPR

GPR is a surface-based active radar system (Daniels, 2004). In order to achieve reasonable penetration depth, frequencies below 1 GHz are typically used in GPR. At a frequency of 1 GHz, penetration depth varies from about 1–2 m (for dry sandy soil or dry rock) to about 10 cm in water-saturated clay. At a frequency of 100 MHz, penetration depth might reach several 100 m (Cook, 1975). Due to its short-range operation, GPR has to radiate very short (of an order of a few ns) microwave pulses into subsurface, which requires ultra-wide operational bandwidth of the system. Commercial GPR systems have typically octave bandwidth (ratio between the highest and lowest operational frequency equals 2), while some experimental systems have this ratio above 10.

While ultra-wide operational bandwidth is one key feature of GPR, another one is ground-coupled antennas. The latter is needed to avoid, as much as possible, reflection of microwaves from air-ground interface and maximize transmitted into the ground microwave power.

GPR nowadays are widely used for a wide scope of applications starting from geological prospecting and ending with landmine detection and classification.

Airborne sensing (CARABAS, radiometers)

Airborne subsurface sensing has been performed both by active (radars) and passive (radiometers) systems. First active systems have been used for polar ice profiling and mapping of central areas of Kalimantan. Furthermore, CARABAS (airborne radar developed by FOA in Sweden and operating at the frequency band 25–85 MHz) has demonstrated capabilities to image buried pipelines in desert conditions (Hellsten et al., 1996), while FOLPEN (radar developed by SRI International and operating at the frequency band 200–400 MHz) has produced images of buried metal-cased antitank landmines in the Yuma Desert (Grosch et al., 1995).

While performance of active radar systems is limited by maximal allowed transmitted power and field breakdown in the antennas, passive systems are free from these limitations. Airborne microwave radiometers have been successfully used in arid regions for search of subsurface

water, water leakage from irrigation channels, and other purposes. Achieving probably similar to active systems penetration depth, passive systems however cannot provide 3D subsurface images as well as comparable to active systems cross-range resolution.

Spaceborne sensing

Similar to airborne systems, satellite-based systems use both radar and radiometer for subsurface sensing. In particular, SIR-A and SIR-C results demonstrated that the L-band radar has penetration capabilities up to 2 m in rigid areas, revealing details of buried rock structures (Elachi et al., 1984).

Conclusion

Having almost 100 years of history, subsurface remote sensing became recently a fast-growing area of science and engineering with steadily expanding range of applications. This subject lies in a border area of three disciplines: microwave interaction with subsurface materials at microscopic level is studied by geophysics; microwave propagation and scattering in lossy, dispersive, and heterogeneous media is studied by electromagnetics; and extraction of information from microwaves radiated or emitted from the subsurface is studied by remote sensing (including GPR). The main research focus in this area remains on interaction of microwaves with heterogeneities of soils and rocks at microscale in order to determine effective dielectric permittivity of matter and development of statistical models for microwave field scattered from subsurface heterogeneities at macroscale.

Bibliography

- Bogorodskii, V., Bentli, C., and Gudmandsen, P., 1983. *Russian Radio Glaciology*. Leningrad: Gidrometeoizdat. In Russian.
- Bojarskii, D. A., Tikhonov, V. V., and Komarova, N. Y., 2002. Model of dielectric constant of bound water in soil for applications of microwave remote sensing. *Progress in electromagnetic Research*, **35**, 251–269.
- Boyarskii, D. A., and Tikhonov, V. V., 1994. Microwave effective permittivity model of media of dielectric particles and applications to dry and wet snow. In *Proceedings of Geoscience and Remote Sensing Symposium*. Vol. 4, p. 2065.
- Cole, K. S., and Cole, H. R., 1941. Dispersion and absorption in dielectrics – I. Alternative current characteristics. *Journal of Chemical Physics*, **9**, 341.
- Cook, J., 1975. Radar transparencies of mine and tunnel rocks. *Geophysics*, **40**, 865.
- Daniels, D. J. (ed.), 2004. *Ground-Penetrating Radar*, 2nd edn. London: The Institution of Electrical Engineers.
- De Loor, G. P., 1983. The dielectric properties of wet materials. *IEEE Transactions on Geoscience and Remote Sensing*, **21**, 364.
- Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., and El-Rayes, M. A., 1985. Microwave dielectric behavior of soil – part II: dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing*, **23**, 35.
- Elachi, C. H., Roth, L. E., and Schaber, G. G., 1984. Spaceborne radar subsurface imaging in hyperarid regions. *IEEE Transactions on Geoscience and Remote Sensing*, **22**, 383.
- Fung, A. K., 1994. *Microwave Scattering and Emission Models and Their Applications*. Norwood: Artech House.
- Grosch, T. O., Lee, C. F., Adams, E. M., Tran, C., Koenig, F., Tom, K., and Vickers, R. S., 1995. Detection of surface and buried mines with an UHF airborne SAR. *Proceedings of SPIE*, **2496**, 110.
- Hallikainen, M. T., Ulaby, F. T., Dobson, M. C., El-Rayes, M. A., and Wu, L. K., 1985. Microwave dielectric behaviour of wet soil – part I. Empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, **15**, 25.
- Heimovaara, T. J., 1994. Frequency-domain analysis of time-domain reflectometry waveforms 1. Measurement of the complex dielectric permittivity. *Water Resources Research*, **30**, 189.
- Heimovaara, T. J., Bouten, W., and Verstraten, J. M., 1994. Frequency domain analysis of time domain reflectometry waveforms 2. A four-component complex dielectric mixing model for soils. *Water Resources Research*, **30**, 201.
- Hellsten, H., Ulander, L. M., Gustavsson, A., and Larsson, B., 1996. Development of VHF CARABAS II SAR. *Proceedings of SPIE*, **2747**, 48.
- Hipp, J. E., 1974. Soil electromagnetic parameters as functions of frequency, soil density and soil moisture. *Proceedings of the IEEE*, **62**, 98.
- Hoekstra, P., and Delaney, A., 1974. Dielectric properties of soils at UHF and microwave frequencies. *Journal of Geophysical Research*, **79**, 1699.
- Hülsenbeck, R., et al., 1926. German patent No. 489434.
- Kovacs, A., Gow, A. J., and Morey, R. M., 1995. The in-situ dielectric constant of polar firm revisited. *Cold Regions Science and Technology*, **23**, 245.
- Olhoeft, G. R., 1987. Electrical properties from 10–3 to 10+9 Hz – physics and chemistry. In Banavar, J. R., Koplik, J., and Winkler, K. W. (eds.), *Physics and Chemistry of Porous Media II*. New York: American Institute of Physics.
- Or, D., and Wraith, J. M., 1999. Temperature effects on soil bulk dielectric permittivity measured by time-domain reflectometry: a physical model. *Water Resources Research*, **35**, 371.
- Sen, P. N., Scala, C., and Cohen, M. H., 1981. A self-similar model for sedimentary rocks with application to the dielectric constant of fused glass beads. *Geophysics*, **46**, 781.
- Shutko, A. M., and Reutov, E. M., 1982. Mixture formulas applied in estimation of dielectric and radiative characteristics of soil and grounds at microwave frequencies. *IEEE Transactions on Geoscience and Remote Sensing*, **20**, 29.
- Simmons, G., Strangway, D. W., Bannister, L., Baker, R., Cubley, D., La Torraca, G., and Watts, R., 1972. The surface electrical properties experiment. In Kopal, Z., and Strangway, D. W. (eds.), *Lunar Geophysics*. Dordrecht: Reidel, p. 258.
- Stern, W., 1929. Versuch einer elektrodynamischen Dickenmessung von Gletschereis. *Gerlands Beiträge zur Geophysik*, **23**, 292.
- Stern, W., 1930. Über Grundlagen, Methodik und bisherige Ergebnisse elektrodynamischer Dickenmessung von Gletschereis. *Zeitschrift Gletscherkunde*, **15**, 24.
- Topp, G. C., Davis, J. L., and Annan, A. P., 1980. Electromagnetic determination of soil water content: measurement in coaxial transmission lines. *Water Resources Research*, **16**, 574.
- Wang, J. R., and Schmugge, T. J., 1980. An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Transactions on Geoscience and Remote Sensing*, **18**, 288.
- Wobischall, D., 1977. A theory of the complex dielectric permittivity of soil containing water: the semidisperse model. *IEEE Transactions on Geoscience and Remote Sensing*, **15**, 49.
- Yarovoy, A. G., de Jongh, R. V., and Ligthart, L. P., 2000. Scattering properties of a statistically rough interface inside a multilayered medium. *Radio Science*, **35**, 455.

MICROWAVE SURFACE SCATTERING AND EMISSION

David R. Lyzenga

College of Engineering, Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, MI, USA

Definition

Surface scattering. The process by which microwave radiation incident upon a solid or liquid surface is wholly or partially redirected away from that surface.

Microwave emission. The process by which microwave radiation originates from a solid or liquid surface due to the rotational or vibrational motions of molecules near the surface.

Introduction

This topic deals with the processes through which a solid or liquid surface influences the electromagnetic field (at microwave frequencies) in the vicinity of that surface and the manner in which these processes are determined by the physical and chemical properties of the surface. The processes of surface scattering and emission, though distinctly different, are related by *Kirchhoff's law* and can both be described by means of a single function, the *bidirectional reflectance*. The bidirectional reflectance function is defined in the following section, and the dependence of this function on the physical and chemical properties of the surface is discussed in the subsequent section. For active microwave (radar and scatterometry) purposes, the scattering properties of surfaces are more commonly described in terms of the normalized radar cross section, or cross section per unit area, which is related to the bidirectional reflectance as discussed near the end of this section.

Bidirectional reflectance

The bidirectional reflectance of a given surface, denoted here by the symbol $\rho(\mu, \phi; \mu', \phi')$, may be loosely described as the proportion of the radiance incident upon the surface from the direction (μ', ϕ') which is scattered or reflected from the surface into the direction (μ, ϕ) . Here μ refers to the cosine of the angle between the direction of the reflected radiation and the surface normal, and ϕ refers to the azimuthal angle of the reflected radiation about the surface normal. The symbols μ' and ϕ' similarly refer to the direction from which the incident radiation arrives at the surface. If we denote the surface normal direction by the unit vector $\mathbf{n}$ and the direction of propagation of the reflected and incident radiation by the unit vectors $\mathbf{k}$ and $\mathbf{k}'$, respectively, then $\mu = \mathbf{k} \cdot \mathbf{n}$ and $\mu' = -\mathbf{k}' \cdot \mathbf{n}$. For a specular surface, all of the incident radiation is reflected into the direction $\mu = \mu'$ and $\phi = \phi' + \pi$, whereas for a diffusely reflecting surface, the radiation incident from a given direction is scattered into a range of angles.

The bidirectional reflectance may be defined more rigorously, albeit implicitly, by the equation

$$L(\mu, \phi) = \int_0^1 \int_0^{2\pi} \rho(\mu, \phi; \mu', \phi') L(\mu', \phi') \mu' d\mu' d\phi' \quad (1)$$

where $L(\mu, \phi)$ is the reflected radiance and $L(\mu', \phi')$ is the incident radiance (Nicodemus, 1965). For a system in thermal equilibrium, the incident radiance is isotropic and is equal to $B_\nu(T_s) = \frac{2}{\lambda^2} \frac{h\nu}{e^{h\nu/kT_s} - 1}$ where $\nu = c/\lambda$ is the frequency, h is Planck's constant, k is Boltzmann's constant, and T_s is the surface temperature. In this case, the sum of the reflected and emitted radiation is also equal to $B_\nu(T_s)$. The emitted radiance in the direction (μ, ϕ) is then equal to $\varepsilon(\mu, \phi) B_\nu(T_s)$, where

$$\varepsilon(\mu, \phi) = 1 - \int_0^1 \int_0^{2\pi} \rho(\mu, \phi; \mu', \phi') \mu' d\mu' d\phi' \quad (2)$$

is the surface emissivity. This relationship is known as Kirchhoff's law for unpolarized radiation. At microwave frequencies, $h\nu \ll kT_s$ for temperatures typical of the Earth's surface, and $B_\nu(T_s) \approx 2kT_s/\lambda^2$ (Rayleigh-Jeans law). Thus, the emitted radiance is proportional to the product of the surface temperature and the emissivity and is often converted into units of temperature (K) and referred to as the *brightness temperature*, this being defined as the temperature of a blackbody that emits the same radiance as the actual surface, i.e., $T_B = \varepsilon T_s$.

Thus far, we have only considered the case of unpolarized radiation. However, the quantities defined above can be extended to the case of polarized radiation as well. There are various definitions of the state of polarization, but for microwave radiation, the most common definition is in terms of the modified *Stokes parameters* T_h, T_v, U , and V (see entry on "*Radiation, Polarization, and Coherence*"). A complete description of the reflectivity would require 16 components, corresponding to each possible combination of incident and reflected polarization states. However, for most remote-sensing applications, we are primarily interested in the reflection of atmospheric (and cosmic) radiation which is virtually unpolarized, and we can combine these into four components corresponding to the reflection of unpolarized radiation into each of the four polarization states. For notational purposes, we indicate that the bidirectional reflectance is polarization dependent by using the bold-faced symbol $\boldsymbol{\rho}(\mu, \phi; \mu', \phi')$, to be understood as a vector with components corresponding to the four modified Stokes parameters. Since the blackbody radiation is unpolarized, the blackbody emissivity vector can be written as $\boldsymbol{\varepsilon}_b = [1, 1, 0, 0]$, and Kirchhoff's law for polarized radiation becomes

$$\boldsymbol{\varepsilon}(\mu, \phi) = \boldsymbol{\varepsilon}_b - \int_0^1 \int_0^{2\pi} \boldsymbol{\rho}(\mu, \phi; \mu', \phi') d\mu' d\phi'. \quad (3)$$

Just above the surface of the Earth, the upwelling radiation field includes contributions from both the downwelling (atmospheric and cosmic) radiation that is reflected from the surface and the radiation that is emitted from the surface. The reflected contribution can be written in terms of the brightness temperature as

$$T_r(\mu, \phi) = \int_0^1 \int_0^{2\pi} \rho(\mu, \phi; \mu', \phi') T_d(\mu', \phi') d\mu' d\phi' \quad (4)$$

where $T_d(\mu', \phi')$ represents the (unpolarized) downwelling radiation at the surface, and the surface-emitted component can be written as

$$\begin{aligned} T_e(\mu, \phi) &= T_S \epsilon(\mu, \phi) \\ &= T_S \epsilon_b - T_S \int_0^1 \int_0^{2\pi} \rho(\mu, \phi; \mu', \phi') d\mu' d\phi'. \end{aligned} \quad (5)$$

Combining these expressions, the total upwelling brightness temperature just above the surface can be written as

$$\begin{aligned} T(\mu, \phi) &= T_e + T_r \\ &= T_S \epsilon_b + \int_0^1 \int_0^{2\pi} \rho(\mu, \phi; \mu', \phi') \\ &\quad [T_d(\mu', \phi') - T_S] d\mu' d\phi'. \end{aligned} \quad (6)$$

Measured at spacecraft altitudes, the observed brightness temperature would of course be reduced by attenuation in the atmosphere and augmented by the upwelling radiation emitted from the atmosphere.

Dependence of bidirectional reflectance on surface properties

The interest in making measurements of the microwave radiation field above the surface of the Earth is to obtain information on geophysical or biological processes occurring at the surface or in the atmosphere or on the physical or chemical state of the surface and/or atmosphere. The types of information and the methods used to extract this information are discussed in other entries (see, e.g., *Land-Atmosphere Interactions*, *Evapotranspiration*; *Ocean-Atmosphere Water Flux and Evaporation*; *Ocean, Measurements and Applications*; *Cryosphere, Measurements and Applications*). In this section, we discuss some of the relationships that underlie these applications.

As discussed in the previous section, the effects of the surface on the radiation field above the surface can be described in principle by the surface temperature and the bidirectional reflectance vector $\rho(\mu, \phi; \mu', \phi')$. This quantity can be defined for a specific surface but more often refers to a statistical ensemble of surfaces with similar properties. It could in principle be measured, but as can

be imagined, this is a difficult and time-consuming process since it involves many different combinations of incidence and reflection angles and must be repeated for many surfaces to obtain a stable statistical mean. Thus it is desirable to model the bidirectional reflectivity when possible, although measurements are necessary for model validation and can sometimes be used directly if some assumptions are made about the directional dependence.

Models have been developed, even for complex surfaces such as vegetation canopies, as discussed in the *Measurements and Applications* entries referenced above. A review of these models is beyond the scope of this entry, but in general the reflectivity depends on the surface roughness, or the geometrical shapes of the elements comprising the surface or canopy, as well as the *dielectric constant* of the component materials. Since the dielectric constant is influenced by such factors as the water content of soils and vegetation, the salinity and phase (liquid or ice) of the ocean surface, and possibly other chemical constituents, the radiation field potentially contains information about these factors. In addition, the roughness of the ocean surface is a function of the wind speed and direction (as well as, possibly, other factors such as the presence of surface slicks, converging or diverging surface currents, and atmospheric stability conditions). As a result, radiometric measurements from aircraft and satellite platforms have been shown to be very useful for measuring surface winds over the ocean on regional and global scales and for observing some of the other features mentioned above on smaller scales. Polarimetric measurements such as those provided by WINDRAD (Yueh et al., 1999) and WindSat (Gaiser et al., 2004) are particularly useful for separating atmospheric from surface effects and for inferring the wind direction as well as wind speed.

For the ocean surface, the bidirectional reflectance is most commonly modeled using a two-scale model (Wentz, 1975; Yueh, 1997) although other models such as the small-slope approximation (Irisov, 1997; Johnson and Zhang, 1999) have also been used. Such modeling also requires a representation of the ocean wave spectrum and its dependence on the wind speed. The spectral model of Durden and Vesecky (1985) has been widely used for this purpose, but some modifications have also been suggested (Lyzenga, 2006) in order to improve comparisons with satellite data. The shortwave (centimeter-scale) portion of the ocean wave spectrum has a strong influence on the microwave surface reflectivity, and there are large uncertainties in this portion of the spectrum due to the difficulty of measuring the surface on these scales. The effects of wave breaking and surface foam on the reflectivity are also incompletely understood, although considerable work has been done in this area.

In the radar literature, the scattering properties of objects are most often described in terms of the radar cross section σ , which is defined as 4π times the reflected power per unit solid angle divided by the power density (power per unit area) incident on the object. It can be defined for any scattering angle but is most commonly used for

backscatter, i.e., for the case in which the source of radiation and the receiver are collocated. This definition applies to discrete objects but was later extended to the case of scattering surfaces by dividing the reflected power by the surface area illuminated or resolved by the radar. This dimensionless quantity is denoted by the symbol σ^o and referred to as the normalized radar cross section, or the cross section per unit area.

The relationship between the radar cross section and the bidirectional reflectance can be found by considering the radiation scattered from a unit surface area. For the case of a distant source, the incident radiance can be written as

$$L(\mu', \phi') = E_o \delta(\mu' - \mu_o) \delta(\phi' - \phi_o) \quad (7)$$

where E_o is the incident power per unit area normal to the direction of the incident radiation (as in the definition of the radar cross section). The scattered radiance for this case is given by Equation 1 as

$$L(\mu, \phi) = \mu_o E_o \rho(\mu, \phi; \mu_o, \phi_o). \quad (8)$$

Following from the definition of the radiance, the reflected power per unit solid angle is equal to $\mu L(\mu, \phi)$. Applying the definition of the radar cross section, we then have

$$\sigma^o = 4\pi \frac{\mu L(\mu, \phi)}{E_o} = 4\pi \mu \mu_o \rho(\mu, \phi; \mu_o, \phi_o) \quad (9)$$

in agreement with Tomiyasu (1988). For the backscatter case, of course, $\mu = \mu_o$.

Conclusion

The effect of the Earth's surface on the naturally occurring microwave radiation field above the surface can be described in terms of the surface temperature and the bidirectional reflectance of the surface. The bidirectional reflectance is a function of the geometry and composition of the surface. For land surfaces, the reflectance is strongly dependent on the water content of the soil and/or vegetation. For water surfaces, the reflectance depends on the salinity, the phase (liquid or solid), and the surface roughness. For the liquid ocean, the surface roughness is primarily dependent on the wind speed but may also be influenced by surface slicks, currents, and atmospheric stability effects.

Bibliography

- Durden, S. P., and Vesecky, J. F., 1985. A physical radar cross-section model for a wind-driven sea with swell. *IEEE Journal of Oceanic Engineering*, **10**, 445–451.
- Gaiser, P. W., St Germaine, K. M., Twarog, E. M., Poe, G. A., Purdy, W., Richardson, D., Grossman, W., Jones, W. L., Spencer, D., Golba, G., Cleveland, J., Choy, L., Bevilacqua, R. M., and Chang, P. S., 2004. The WindSat spaceborne polarimetric microwave radiometer: sensor description and early orbit performance. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 2347–2361.

- Irisov, V. G., 1997. Small-slope expansion for thermal and reflected radiation from a rough surface. *Waves in Random Media*, **7**, 1–10.
- Johnson, J. T., and Zhang, M., 1999. Theoretical study of the small slope approximation for ocean polarimetric thermal emission. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 2305–2316.
- Lyzenga, D. R., 2006. Comparison of WindSat brightness temperatures with two-scale model predictions. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 549–559.
- Nicodemus, F. E., 1965. Directional reflectance and emissivity of an opaque surface. *Applied Optics*, **4**, 767–773.
- Tomiyasu, K., 1988. Relationship between and measurement of differential scattering coefficient (σ^o) and bidirectional reflectance distribution function (BDRF). *IEEE Transactions on Geoscience and Remote Sensing*, **26**, 660–665.
- Wentz, F. J., 1975. A two-scale model for foam-free sea microwave brightness temperatures. *Journal of Geophysical Research*, **80**, 3441–3446.
- Yueh, S. H., 1997. Modeling of wind direction signals in polarimetric sea surface brightness temperatures. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, 1400–1418.
- Yueh, S. H., Wilson, W. J., Dinardo, S. J., and Li, F. K., 1999. Polarimetric microwave brightness signatures of ocean wind directions. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 949–959.

Cross-references

[Cryosphere, Measurements and Applications](#)
[Land-Atmosphere Interactions, Evapotranspiration](#)
[Ocean-Atmosphere Water Flux and Evaporation](#)
[Ocean, Measurements and Applications](#)
[Ocean Measurements and Applications, Ocean Color](#)
[Radiation, Polarization, and Coherence](#)

MISSION COSTS OF EARTH-OBSERVING SATELLITES

Randall Friedl and Stacey Boland
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Synonyms

Earth-observing mission costs; Spaceborne mission costs

Definition

Earth-observing satellite mission. An instrumented spacecraft deployed in space that is designed to observe select geophysical parameters of the Earth for purposes of environmental science or monitoring. A satellite mission requires a launch vehicle, observing instruments, and a spacecraft bus which houses the instruments as well as systems for guidance, navigation, control, communications, command and data, power, thermal control, propulsion, and structures.

Satellite mission cost. The total cost of implementing a satellite mission. The cost includes all of the component

costs as well as costs for design, system engineering, operating and data analysis software, and mission operations. The cost does not typically include scientific analysis of the data collected by the mission.

Introduction

Observing the Earth from the vantage point of space has transformed our understanding of the Earth as an integrated system of physical and biological components. The number of Earth-observing satellites has steadily increased with time, with more than 100 satellite missions currently operating. The desire for new satellite missions remains high, especially since many of the geophysical parameters of interest to scientists (e.g., 44 essential climate variables as defined by the Global Climate Observing System Program) are still not observed from space and other key parameters need to be sustained over many satellite lifetimes to enable investigations of long-term climate trends. The need for new missions is tempered by the relatively high cost, currently ranging from \$100 M to \$3 B, to deploy and operate a satellite mission. Understanding the factors that determine the costs of satellite missions is critical to developing and sustaining an effective system of Earth-observing satellites.

Cost factors

Transporting instrumented science payloads to space requires powerful launch vehicles. The relatively small number (i.e., ~30 total, ~12 USA) of available launch vehicles is divided into four classes, as defined by the US Federal Aviation Administration Office of Commercial Space Transportation, that reflect lift capability as follows: small ($\leq 5,000$ lb to Low Earth Orbit or LEO), medium (5,001–12,000 lb to LEO), intermediate (12,001–25,000 lb to LEO), and heavy ($\geq 25,000$ lb to LEO). Examples of launch vehicles across the four classes are Pegasus (small), Delta II (medium), Atlas IIAS (intermediate), and Atlas V (heavy). While the cost of a launch vehicle increases with each step-up in class, the per pound payload cost is roughly the same across classes when the launch vehicle lift capability is fully utilized. There is thus a substantial cost incentive to either design a payload to fit into the smallest possible launch vehicle or to combine payloads to take advantage of the full capability of a larger-class vehicle. Most of the past and current Earth science missions range between 1,000 and 10,000 lb (instruments plus spacecraft) and were launched on small- and medium-class launch vehicles.

While launch vehicle costs can be directly correlated to performance requirements in terms of a “price per pound” metric, this is not the case for estimating the cost of the satellite instruments. Science instrument payloads typically represent the most unique and challenging components of Earth-observing missions. Payload designs are generally mission unique, resulting from a careful optimization of science requirements against physical and programmatic constraints (e.g., cost, schedule, risk). Instrument

costs can thus be driven by a variety of factors. Increasingly stringent requirements on temporal and spatial resolution, for example, might necessitate larger signal collection optics and/or antennas as well as highly sensitive detectors and precision signal processing electronics, driving payload mass, volume, and component costs higher. Attempts to reduce payload mass and volume might necessitate incorporation of new, yet expensive, lightweight technologies.

Spacecraft costs, like launch vehicle costs, are generally performance driven. Instrument accommodation requirements, particularly mass, power, data rate, and configuration constraints, greatly dictate requirements for the platform, or spacecraft bus, that houses the instrument. While many standardized platforms are now available, most missions have the need for some degree of spacecraft customization (e.g., tighter thermal control, improved pointing accuracy or knowledge, larger onboard data storage capacity, additional power allocations). The cost of customization is to a large extent dependent on whether new technology development is required.

Mission operations costs are driven, to a large extent, by mission data volume and latency requirements which can drive ground network, data processing, and archival costs. The spacecraft’s onboard data storage capacity (i.e., solid-state recorder size) is also a determining factor in the required frequency of data downlinks and/or number of ground stations.

Cost estimation methods

Three methodologies are generally used to estimate mission costs, namely, analogy, parametric, and “grass roots.”

An analogy-based estimate represents the simplest approach to cost estimation. It relies on direct comparisons to analogous ongoing or previous missions. Provided there is little deviation from the chosen analog, rapid and accurate mission cost estimates can be obtained using this method. However, since very few missions share identical subsystems or components, analogy-based estimates are not widely applicable without including cost adjustments based on analysis of differences between the target mission and the analog. Such adjustments typically employ parametric analyses in order to enable quantitative extrapolations between systems with varying degrees of differences. Analogy-based estimates best apply to mission lines which involve launching a series of similar (or near-identical) satellites (e.g., the NOAA POES series of weather satellites) and those which incorporate incremental performance improvements between generations (e.g., the NASA TOPEX-Poseidon, Jason, OSTM-sustained research satellite series).

Parametric cost estimates are based on cost driver relationships derived from analyses of historical data. The primary cost drivers used for parametric studies typically include mass, power, and subsystem complexity. For example, experience shows that mission costs increase roughly linearly with mass within a given launch vehicle

class, allowing derivation of the “price per pound” metric cited above. Once developed, parametric relationships are particularly useful for performing rapid trade studies during the initial scoping of mission concepts, particularly for concepts without direct historical analogy.

Grassroots cost estimates involve detailed consideration of each element of a mission or component of an instrument. For each element or component, a cost estimate is generated and basis of estimate is documented, describing the technique used to derive the estimate. For a true grassroots estimate to be performed, the mission or instrument design must be specified to a fairly high degree to enable rigorous quantitative cost estimates based on material and labor costs allocated over a defined implementation schedule. Consequently, grassroots estimates find their greatest utility during later stages of mission development. During concept development, a “quasi-grassroots” mechanism can be employed in which the grassroots cost estimating structure is preserved, including documentation of a basis of estimate, while leveraging analogy or parametric cost estimates for the individual components for which detailed requirements, designs, and/or schedules are not well known.

Cost estimation accuracy

The ability to conceive, plan, and implement an effective and affordable multi-satellite observing system rests in large part on understanding the costs of individual missions. In particular, decisions regarding mission priorities and sequencing are intimately tied to considerations of cost to benefit.

Because there are relatively small numbers of satellite missions and most of them are partially, or wholly, unique in design, the ability to accurately estimate cost is limited in the beginning phases of a mission. Analysis of historical mission cost data reveals a tendency for substantial increases (10–50 %) in estimates for lower-cost Earth missions when going from initial to later mission phases. These increases have been traced primarily to initial underestimations of instrument and spacecraft components. In many cases, the cost increases reflect more stringent measurement requirements stemming from refinement of the mission science goals, rather than inherent inaccuracies in the estimation methodology.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Bearden, D., 2005. Perspectives on NASA mission cost and schedule performance trends. In *NASA Goddard Flight Center Symposium*.
- Committee on Earth Observation Satellites mission database, available at <http://www.ceos.org/>
- Emmons, D. L., Bitten, R. E., and Freaner, C. W., 2006. Using historical NASA cost and Schedule growth to set future

program and project reserve guidelines, IEEEAC paper #1545, Version 3.

FAA Commercial Launch Vehicle website. http://www.faa.gov/about/office_org/headquarters_offices/ast/launch_data/.

Futron Corporation, 2002. Space transportation costs: trends in price per pound to orbit 1990–2000. www.futron.com.

Global climate observing system program’s essential climate variables. Available from <http://www.wmo.int/pages/prog/gcos/index.php?name=essentialvariables>.

Larson, W. J., and Wertz, J. R., 1999. *Space Mission Analysis and Design*. El Segundo: Microcosm, Vol. 3.

NASA, 2008. Cost estimation handbook. Available from <http://ceh.nasa.gov>.

Wertz, J. R., and Larson, W. J., 2007. *Reducing Space Mission Cost*. New York: Springer.

Cross-references

[Global Climate Observing System](#)
[Mission Operations, Science Applications/Requirements](#)
[Observational Systems, Satellite](#)

MISSION OPERATIONS, SCIENCE APPLICATIONS/REQUIREMENTS

David L. Glackin
 Los Angeles, CA, USA

Synonyms

Environmental data record (EDR) requirements

Definition

Science applications. Environmental phenomena to be measured in the Earth’s atmosphere, oceans, land surface, solid Earth, ice cover, and near-Earth space environment. *Science requirements.* A numerical specification of the phenomena that must be measured and a corresponding set of attributes including spatial resolution, spatial reporting interval, spatial coverage, measurement uncertainty (accuracy/precision), long-term measurement stability, mapping uncertainty, reporting frequency, and timeliness of data delivery.

Introduction

The science applications are described thoroughly elsewhere in this volume (see *Atmospheric General Circulation Models*; *Ocean, Measurements and Applications*; *Cryosphere, Measurements and Applications*; *Land-Atmosphere Interactions, Evapotranspiration*; *Trace Gases, Troposphere - Detection from Space*). The thrust of this entry is the science requirements and how they are motivated, established, codified, analyzed, and iterated.

David L. Glackin: deceased.

Motivation for and establishment of requirements

In an ideal world, science requirements should be motivated by a problem in need of a remote sensing instrument (or set of instruments), not vice versa. Reality is sometimes different, but environmental remote sensing missions, especially large ones, typically have a large user constituency behind them that provides the rationale for the mission. Mission requirements may be driven by the need for new types of data, improvements in the quality of old types of data, or continuity of data for studies of long-term trends.

Ideally, the requirements flow should be from the top down. The science requirements should drive the instrument concepts. Those concepts in turn should drive the instrument designs, which should in turn drive the platform (satellite, aircraft, etc.) concept and design. Examples of important aspects of the platform design that the science requirements drive are pointing accuracy and stability and the choice of satellite orbit (see *Observational Systems, Satellite*).

As an adjunct to the science requirements, the postlaunch data reduction, analysis, and research activities should be considered up front. Too often, the tendency is to focus on building the instruments and the platform to meet the requirements and to delay consideration of the analysis and research. The latter approach can have large systems impacts, and it is best to consider these issues up front while setting requirements. For example, the calibration/validation phase of a mission is a major postlaunch activity and is the one in which the instrument calibration is established and in which the quality of the science data products is validated (see *Calibration and Validation*). If the science requirements lead to data products that are exceedingly difficult to validate, unexpected amounts of postlaunch resources may be required (see *Mission Costs of Earth-Observing Satellites*).

Science requirements are usually established by user working groups, which are organized by scientific discipline. It is important to involve the instrument systems engineers early in this process. They understand the hardware in an end-to-end systems sense and can help the user communities to understand the cost and risk implications of their requirements. Often, scientific “desires” must be tempered with technological reality. Sometimes, it may be possible to achieve half of what is desired at one-tenth the cost. Science requirements are typically iterated between these two groups of people until a realistic set of requirements is established.

Science requirements are often couched in terms of “thresholds” and “objectives.” The thresholds are the minimum set of requirements that must be met. The objectives represent the performance that the users would like to meet, if meeting them does not inordinately drive the hardware design. Remote sensing missions may in the end deliver data products that fall between these two benchmarks.

Codification of requirements

The science requirements must be established carefully. Clarity is of the utmost importance. If the requirements

are not clear up front, much time will be wasted later in clarifying the requirements to the instrument and platform vendors and then in rewriting the requirements. Definitions must be clear and comprehensive. Units must be consistent and understandable. Requirements must be consistent with basic physics. There should be no conflicting requirements. Very significant is the establishment of requirements for measurement quality. It is possible to generate major confusion over items as seemingly simple as measurement accuracy, precision, and uncertainty. The latter can be a major stumbling block in the process of requirements establishment and should be treated thoughtfully.

Requirements analysis and iteration

Science requirements should be thoroughly analyzed after they are initially established, for clarity, consistency, comprehensiveness, lack of conflict, and impact on the end-to-end mission. This is the point at which the design engineers, the hardware vendors, and the operators should engage with the system engineers to ensure that the set of requirements is sound. At this stage, iterations of the requirements are usually made, in consultation with the user community.

It is paramount that experienced people think through these issues up front. Otherwise, problems can crop up during the course of the program that can be very wasteful of resources.

It is particularly important to define the “driving requirements” (Wertz and Larson, 1991). These are the ones that dictate some of the major hardware characteristics.

“Requirements creep” is almost always a problem in any large remote sensing mission. Once the user communities start to see what the system can do, they usually think of more things that it might do. One solution for that is to define these items as “Pre-Planned Product Improvements” (P3I) that might be addressed later, given sufficient resources.

Conclusion

An intelligent process of establishing the science requirements for a mission can make the difference between mission success and failure or between delivering the hardware on time and under budget and delivering it late and far over budget. Getting the science requirements right at the beginning and thinking things through thoroughly up front is key to mission success.

Bibliography

Wertz, J. R., and Larson, W. J., 1991. *Space Mission Analysis and Design*. Dordrecht: Kluwer.

Cross-references

[Calibration and Validation](#)
[Cryosphere, Measurements and Applications](#)
[Ocean, Measurements and Applications](#)

O

OBSERVATIONAL PLATFORMS, AIRCRAFT, AND UAVS

Jeffrey Myers
NASA/Ames Research Center, Airborne Science and
Technology Laboratory, University of California,
Santa Cruz, Moffett Field, CA, USA

Synonyms

Airborne platforms; Remote sensing aircraft; Suborbital platforms

Definition

Airborne observational platforms are any aircraft, manned or unmanned, which is equipped to obtain remotely sensed data or in situ atmospheric samples above the surface of the Earth.

Introduction

Airborne platforms are typically used to obtain observations of high spatial or temporal resolution that cannot be practically achieved with satellite- or ground-based systems. These include remote sensing observations of dynamically evolving events (e.g., volcanic eruptions or convective weather systems) which could not be made by satellites due to their fixed revisit schedules or in situ air sampling of atmospheric constituents at varying altitudes. Because they can selectively operate during clear-sky conditions, they are also often used for remote sensing in areas which have persistent cloud cover. Imagery in some tropical, mountainous, and arctic regions, for example, cannot be obtained from polar-orbiting satellites on any regular basis.

The spatial resolution of imagery obtained from an airborne platform can be precisely controlled by varying the altitude of the aircraft above the ground. Pixel sizes

produced from a sensor with a 2.5 mrad instantaneous field of view (IFOV), for example, can thus be varied between 3 and 50 m with the same instrument, when operated at altitudes of 1,200 and 19,800 m, respectively. This variability allows the imagery to be tailored to the requirements of the specific phenomenon being studied. Most satellite systems, especially those with the infrared spectral bands of interest to earth scientists, are designed for broadscale global observations at fixed spatial resolutions, which may be too coarse for many detailed geophysical process studies. Spatial resolutions of airborne infrared imaging systems can exceed 3 m and can be as high as 1 cm in the visible and near-infrared wavelengths, depending on platform altitude.

Airborne platforms are also used as test beds for new instrument development. They provide an economical method of testing prototype satellite sensors prior to launch and also for collecting the empirical data needed to develop the associated retrieval algorithms for new orbital systems. Once a satellite system is placed in orbit, airborne sensors are often used to conduct simultaneous, high-resolution observations for data validation and instrument calibration purposes.

Overview of platform types

Manned aircraft

A wide variety of airborne platforms are used for Earth observations, ranging from small single-engine propeller planes with minimal modifications to large specially adapted multiengine commercial and military reconnaissance aircraft. Unmanned Aerial Vehicles (UAVs), airships, and balloons also fill unique observational niches. These platforms are modified to varying degrees to accommodate either remote sensing or air sampling payloads. The particular measurement requirements, together with budget constraints, indicate the most appropriate platform for a given application. Fundamental aircraft

variables to be considered include operating range and altitude, payload capacity, flight endurance, and cost of operation. There is typically a trade-off with most aircraft between the weight of the payload and the amount of fuel to be carried, which in turn affects the operating range.

If the instruments to be flown are relatively lightweight and the projected operating area is small and close to an airfield, a light one- or two-engine private plane is by far the most economical. Typically unpressurized, they usually operate at altitudes below 7,500 m with ranges under 600 km. As the size of the instrument payload and/or the extent of the study area increases, larger and more capable platforms are required. Turbine or turbo-charged twin engine propeller aircraft offers the next level of performance with a significant increase in payload and range, pressurized cabins, and operating altitudes up to 10,000 m. These often offer the best compromise between performance and cost-effectiveness for many applications. Corporate jets or larger commercial transport aircraft, which can fly above 12,000 m and at much higher speeds, can carry heavy payloads and cover even larger areas in a single flight but are relatively expensive to operate. Special-purpose high-altitude platforms, such as the NASA ER-2 (a Lockheed U-2 reconnaissance aircraft derivative) or the Russian M-55 Geophysica, can fly into the lower stratosphere and operate over very large geographical areas. These aircraft are often used for atmospheric research but are still more expensive and require specialized ground support equipment. For remote sensing missions, the very long atmospheric path length below these aircraft makes the atmospheric correction task more problematic. Imaging through a nearly full atmospheric column can be an advantage for the validation of satellite data, however. Flying at 20,000 m altitude, these aircraft are above 95 % of the Earth's atmosphere, allowing for very accurate at-sensor response comparisons between airborne and orbital systems.

UAVs: unmanned aerial vehicles

Also known as Unmanned Aerial Systems (UAS), these include a full range of size classes from very small hand-launched aircraft resembling model airplanes to the large high-altitude observational systems used by military services and some science agencies. The complexities of operation, together with payload and range capabilities, also vary conversely. Most UAVs are remotely piloted by a person on the ground, using a radio link to send commands to the flight control system on the aircraft. However, some are configured to operate autonomously, using an onboard computer with a preprogrammed flight plan, together with a satellite Global Positioning System (GPS) and an inertial guidance system. UAVs offer several significant advantages over manned platforms. For example, they can operate in hazardous environments (e.g., near volcanic activity or in toxic plumes) without endangering a human pilot. Some UAVs also have endurance in excess of 20 h, which enables sustained observations of

an evolving phenomenon. Airships, both rigid dirigibles and nonrigid types, also have this capability.

Apart from vehicles designed for completely autonomous operation, some form of communication link is generally required to operate a UAV. This can consist of a simple line-of-sight radio link as used on small model airplanes, which limits the range of operation, or a larger satellite communication ("satcom") system, which extends the operating range over the horizon. The most elementary satellite communications systems use mobile satellite telephone technology with small omnidirectional antennas and conventional low-speed data modems. The higher data-rate satcom systems found on larger UAVs require large directional antennas and are significantly heavier and more expensive to operate.

The smaller UAVs can be very economical overall to operate, while the larger military-type platforms, which use high-speed satellite communications links and complex ground stations with professional pilots, can be considerably more expensive. Regardless of size, however, almost all UAVs are subject to some form of government air traffic control regulations, which can make flight operations – especially over urban areas – highly problematic. In the USA, even the smallest model airplane types are subject to federal airspace restrictions when used for any purpose other than recreation. Access to the local airspace for a UAV should never be assumed and indeed has been a major obstacle to their wider proliferation in the private and scientific sectors.

Payload accommodations

Airborne observation platforms require varying degrees of mechanical modification to suit their intended purpose. These can range from simple external instrument mounting fixtures on an otherwise unmodified light aircraft to permanent airframe modifications that may include optical windows, air sampling probes, or external attachment points for sensor pods.

Remote sensing aircraft

These aircraft are usually fitted with optical camera windows at various angles for nadir, side, or zenith viewing or external pods with similar accommodation. If the windows are fitted in an aircraft with a pressurized cabin, they must be specially engineered with adequate safety margins. Windows in an unpressurized aircraft (typically operating at altitudes less than 3,000 m) are less problematic to install. Commercially available aircraft optical windows typically have wavelength transmissions between 180 nm and 2.5 μm . Infrared windows are also available, although generally in smaller sizes, and are much more expensive. Many infrared instruments instead are designed to view through an open aperture in the aircraft fuselage to avoid window cost and/or spectral transmission issues. In the case of a pressurized aircraft, this open observation port, together with the instrument itself, is enclosed beneath a sealed housing which preserves the

aircraft cabin pressurization. These installations often include an external sliding door that covers the open aperture to protect the instrument when not in use.

Airborne radar platforms

Airborne radar or passive microwave systems require various forms of antennas, which are either integral to the airframe or carried in external pods. Flat-panel phased-array antennas which have no moving parts can be affixed directly to the outside of the aircraft. Mechanically pointed antennas are typically mounted under domes or in external pods made of a material that is transparent to the relevant frequencies. Depending on the particular system, these antennas can be quite large and may require extensive modifications to the platform. They may also have large electrical power requirements, which can require special modifications to the aircraft power system.

Air sampling platforms

These platforms are equipped with inlets to bring the external ambient air into measurement devices mounted either inside the cabin or in externally mounted pods. These usually consist of external probes or booms on the aircraft nose, fuselage, or wings, which extend the inlets into the free airstream. The inlet probes must extend beyond the boundary layer of the platform and be otherwise free of any turbulence or aerodynamic effects from the aircraft itself, to ensure unbiased sampling of free air at ambient pressure. The location of the probes on the aircraft is usually determined by modeling the airflow around the particular fuselage, under the airspeed and attitude parameters (pitch, roll, yaw) expected during sampling operations.

Data systems

Observation platforms are also often fitted with ancillary data systems to augment the primary instrument measurements. These may include the basic recording of aircraft state parameters, such as time, navigational position and attitude, altitude, and heading, as well as external air temperatures and pressures. These data are usually provided by some combination of satellite GPS and the aircraft's own Inertial Navigation System (INS) or Air Data Computer. For applications requiring very precise knowledge of platform attitude and location, such as the precision geo-location of image data, or for quantitative air sampling, stand-alone Inertial Measurement Unit (IMU) and differential GPS units may be used. In the air sampling case, this may be coupled with external pressure transducers, which allows for the derivation of external winds around the platform.

Other considerations

The operating environment for instruments onboard an aircraft is typically demanding and may impose stresses on the equipment that are not always conducive to good

scientific measurement. Outside air temperatures vary radically with altitude, often causing instruments to cool considerably unless mounted inside a pressurized cabin. Those with thermal infrared channels viewing through an open port are particularly subject to severe environmental stresses, including temperature cycling, water condensation, and optical contamination. The use of heaters to maintain the temperature of critical electrical and optical components above the local dew point, and hermetic sealing of optics and spectrometers wherever possible, has been proven to increase data quality as well as the longevity of the systems themselves. The deposition of contaminants onto glass surfaces (lenses, windows, etc.) can also compromise measurement quality, so regular cleaning of these surfaces is necessary. The electrical power found on many commercial aircraft, usually 28 V Direct Current, together with the associated electrical grounding circuitry, can be highly variable in quality. Instruments should incorporate internal electrical power conditioning whenever possible to mitigate adverse effects.

Mechanical vibration is another potential problem source for delicate instrumentation and each platform has its own vibration frequency spectrum which changes with engine power and control surface settings. Excessive vibration can affect the optical alignment of spectrometers, induce microphonic noise in infrared detectors, and cause the failure of electrical connections. The G-forces associated with landings and in-flight turbulence can also be substantial. Some form of vibration isolation between instrument and airframe, tailored to the weight of the instrument and the specific vibration regime of the platform, is often indicated. In general, instrumentation should be ruggedized to obtain consistent results in the airborne environment.

Summary

Airborne observational platforms consist of a variety of types, including conventional and unmanned aircraft, and have a broad range of technical capabilities and operating costs. Aircraft used for Earth observing require varying degrees of modification, depending on their intended applications. This entry provides a brief overview of platform types, mission-specific modifications, and their relevant operating envelopes.

Bibliography

- Cox, T., Nagy, C., Skoog, M., and Somers, I., 2004. Civil UAV capability assessment. Suborbital Science Program Internal Document. NASA.
- Fladeland, M., and Schoenung, S., 2007. NASA Earth science requirements for suborbital observations. Airborne Science Program Internal Document. NASA.
- Henderson, F. M., and Lewis, A. J. (eds.), 1998. *Manual of Remote Sensing*. New York: Wiley. Principles & Applications of Imaging Radar, Vol. 2.
- Jackson, M. (ed.), 2009. *Manual of Remote Sensing*. New York: Wiley. Earth Observing Platforms & Sensors, Vol. 1.1.

- King, M., Menzel, W. P., Grant, P. S., Myers, J. S., Arnold, G. T., Platnick, S. E., Gumley, L. E., Tsay, S. C., Moeller, C. C., Fitzgerald, M., Brown, K. S., and Osterwisch, F. G., 1996. Airborne scanning spectrometer for remote sensing of cloud, aerosol, water vapor, and surface properties. *Journal of Atmospheric and Oceanic Technology*, **13**(4), 777–794.
- Kramer, H. J., 1996. *Observation of the Earth and Its Environment*, 3rd edn. New York: Springer.
- McDonnell Aircraft Company, Marketing Division, 1982. *Reconnaissance Handy Book*. St. Louis: McDonnell Aircraft.
- Miller, R., Del Castillo, C., and McKee, B. (eds.), 2005. *Remote Sensing of Coastal Aquatic Environments*. Dordrecht: Springer.
- Pepi, J. W., 1994. Fail-safe design of an all BK-7 glass aircraft window. In *Window and Dome Technologies and Materials IV*, SPIE, Vol. 2286, September 28, 1994.
- Rencz, A. N., and Ryerson, R. A. (eds.), 1999. *Manual of Remote Sensing*. New York: Wiley. Remote Sensing for the Earth Sciences, Vol. 3.
- Stefanutti, L., and Sokolov, L., 1999. The M-55 geophysics as a platform for the airborne polar experiment. *Journal of Atmospheric and Oceanic Technology*, **16**(10), 1303–1312.
- Ustin, S., 1999. *Manual of Remote Sensing*. New York: Wiley. Remote Sensing for Natural Resource Management and Environmental Monitoring, Vol. 4.

Cross-references

[Remote Sensing, Physics and Techniques](#)

OBSERVATIONAL SYSTEMS, SATELLITE

David L. Glackin
Los Angeles, CA, USA

Acronyms

ADEOS	Advanced Earth Observing Satellite
ADM	Atmospheric Dynamics Mission
AIM	Aeronomy of Ice in the Mesosphere
ALISSA	l'Atmosphere par Lidar Sur SAliout
ALOS	Advanced Land Observing Satellite
AMI	Active Microwave Instrument
ASCAT	Advanced Scatterometer
ATSB	Astronautic Technology Sdn Bhd (Malaysia)
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CAST	Chinese Association for Science and Technology
CBERS	China-Brazil Earth Resources Satellite
CERES	Clouds and the Earth's Radiant Energy System
CHAMP	CHAllenging Minisatellite Payload
CHRIS	Compact High-Resolution Imaging Spectrometer

David L. Glackin: deceased.

CNES	Centre National d'Etudes Spatiales
COMS	Communications, Oceans and Meteorology Satellite
CONAE	National Commission for Space Activities (Argentina)
COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate
COSMO	Constellation of small Satellites for Mediterranean basin Observation
CSA	Canadian Space Agency
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DARA	German Space Agency
DLR	German Aerospace Center
DMC	Disaster Monitoring Constellation
DMSP	Defense Meteorological Satellite Program
DoD	Department of Defense
EADS	European Aeronautic and Space Company
EarthCARE	Earth Clouds, Aerosols and Radiation Explorer
EO-1	Earth Observer-1
EOS	Earth Observing System
ERBS	Earth Radiation Budget Satellite
EROS	Earth Resources Observation Satellite
ERS	European Remote Sensing
ESA	European Space Agency
ETM+	Enhanced Thematic Mapper Plus
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FEDSAT	Federation Satellite
FY	Feng Yun
GCOM	Global Change Observation Mission
GEO	Geostationary Earth Orbit
GER	Geophysical and Environmental Research Corporation
GERB	Geostationary Earth Radiation Budget
GFZ	German Research Centre for Geosciences
GISTDA	Geo-Informatics and Space Technology Development Agency (Thailand)
GLM	GEO Lightning Mapper
GMES	Global Monitoring for Environment and Security
GOCE	Gravity field and steady-state Ocean Circulation Explorer
GOES	Geostationary Operational Environmental Satellite
GOMS	Geostationary Orbit Meteorological Satellite
GPM	Global Precipitation Measurement
GRACE	Gravity Recovery and Climate Experiment
HJ	Huan Jing
HY	Hai Yang
ICESat	Ice, Cloud, and land Elevation Satellite
IJPS	Initial Joint Polar System
INPE	National Institute of Space Research (Brazil)
IRS	Indian Remote Sensing
ISRO	Indian Space Research Organisation

JAXA	Japanese Aerospace Exploration Agency
JPL	Jet Propulsion Laboratory
KARI	Korea Aerospace Research Institute
KOMPSAT	Korea Multi-Purpose Satellite
LDCM	Landsat Data Continuity Mission
LEO	Low Earth Orbit
LIS	Lightning Imaging Sensor
LITE	Lidar In-space Technology Experiment
MADRAS	Microwave Analysis and Detection of Rain and Atmospheric Structures
MDA	MacDonald, Dettwiler and Associates, Ltd.
MEO	Medium Earth Orbit
MetOp	Meteorological Operational
MIS	Microwave Imager/Sounder
MOMS	Modular Optoelectronic Multispectral Scanner
MSG	Meteosat Second Generation
MTG	Meteosat Third Generation
NASA	National Aeronautics and Space Administration
NMP	New Millennium Program
NOAA	National Oceanic and Atmospheric Administration
NPOESS	National Polar-orbiting Operational Environmental Satellite System
NPP	NPOESS Preparatory Project
NRL	Naval Research Laboratory
NSPO	National Space Organization (Taiwan)
NTU	Nanyang Technological University (Singapore)
OCO	Orbiting Carbon Observatory
OLI	Operational Land Imager
OMI	Ozone Monitoring Instrument
OSTM	Ocean Surface Topography Mission
PARASOL	Polarization and Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar
POES	Polar-orbiting Operational Environmental Satellite
PROBA	Project for On-Board Autonomy
RDL	Research and Development Laboratory
RISAT	Radar Imaging Satellite
ROCSAT	Republic of China Satellite
RSSS	Remote Sensing Satellite System
SAC	Satelite de Aplicaciones Cientificas
SAOCOM	Satellites for Observation and Communication
SAR	Synthetic Aperture Radar
ScaRab	Scanner for Radiation Budget
SeaWiFS	Sea-viewing Wide-Field Sensor
SIR	Shuttle Imaging Radar
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture Ocean Salinity
SNSB	Swedish National Space Board
SPIN-2	Space Information-2
SPOT	Système Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
SSTL	Surrey Satellite Technology Ltd.

THEOS	Thailand Earth Observation System
TOMS	Total Ozone Mapping Spectrometer
TOPEX	TOPOgraphy EXperiment for Ocean Circulation
TPFO	TOPEX/Poseidon Follow-On
TRMM	Tropical Rainfall Measuring Mission
TUBITAK	The Scientific and Technological Research Council of Turkey
UAV	Uninhabited Aerial Vehicle
UCAR	University Corporation for Atmospheric Research
USGS	United States Geological Survey
ZY	Zi Yuan

Synonyms

Earth observation satellites; Earth observation systems; Remote sensing satellites; Remote sensing systems

Definition

Satellite Observational System. A system for the study of the Earth and its environment (atmosphere, oceans, ice, land surface, solid Earth, and near-Earth space environment) from space. The end-to-end system includes (1) the spacecraft (the platform that hosts the remote sensing instruments and provides them with pointing, power, commands, data storage, etc.), (2) the remote sensing instruments (often called “sensors”) that observe and measure the characteristics of the Earth’s environment, and (3) the ground-based mission operations. The term “satellite” is usually used for the combination of the spacecraft and instruments, but in some communities, the term “spacecraft” is used for that combination, with the term “satellite” indicating the platform.

Introduction to satellite observational systems

Although the first weather satellite, TIROS I, was launched in 1960, the field of satellite-based remote sensing of Earth really began to take form in the 1970s (see *Remote Sensing, Historical Perspective*). The launches of Landsat-1 in 1972, Skylab in 1973, Nimbus-7 in 1978, and Seasat in 1978 set the stage for modern environmental remote sensing (Glackin, 2004).

Since the pioneering work of the 1970s, the field of satellite environmental remote sensing has steadily evolved. Before 1990, only six nations owned environmental satellites (China, France, India, Japan, the USA, and the former USSR), but since then, the number has quintupled. The new nations (Glackin and Peltzer, 1999; Kramer, 2002; eoPortal, 2008) include the ones in Africa (Algeria, Morocco, and Nigeria), Asia (Indonesia, Malaysia, South Korea, Taiwan, Thailand, and Ukraine), Europe (Belgium, Germany, Italy, Portugal, and the UK), the Middle East (Egypt, Iran, Israel, Pakistan, and Turkey), North America (Canada), Scandinavia (Sweden), and South America

(Argentina, Brazil, and Chile), as well as the country of Australia. Many of these nations have procured their satellites from other nations, as described below.

Pre-1990 programs primarily involved systems of high cost and complexity, but post-1990 the focus shifted to include missions involving smaller satellites of lower complexity and cost and greater commercial involvement, including hybrid government/commercial systems (see *Public-Private Partnerships*) and fully commercial systems (see *Commercial Remote Sensing*). Examples of hybrid government/commercial systems are the US Landsat, the French SPOT (Système Pour l'Observation de la Terre), and the Canadian RADARSAT satellites. In such hybrid programs, industry is expected to share a portion of the development costs and to recoup their investment by developing a commercial market for the data products (see *Cost Benefit Assessment*).

Fully commercial systems have met with some success, although many efforts have been canceled. These include an Australian hyperspectral (see *Remote Sensing, Physics and Techniques*) mining satellite called ARIES, satellites for precision agriculture called GEROS from Geophysical and Environmental Research Corp. (GER) and Resource21 from Boeing/GDE Systems, the "One-Meter System" from GDE Systems, the Radar-1 SAR (Synthetic Aperture Radar) system from Research and Development Laboratory (RDL), and AVStar, a system from AstroVision to provide color imagery of the Earth from GEO (Geostationary Earth Orbit) (see Glackin and Peltzer, 1999.) The five latter systems were all from the USA.

UoSAT-5, launched by the British company Surrey Satellite Technology Ltd (SSTL; SSTL, 2008) in 1991, was the first commercial microsat. OrbView-1, launched by the US company ORBIMAGE (now part of GeoEye) in 1995, was the first commercial weather satellite (Glackin and Peltzer, 1999). In the 1990s, governments also pioneered the concept of buying end-user data products rather than satellites. The first example involved a NASA (the National Aeronautics and Space Administration) data buy of data from OrbView-2/SeaStar, the first satellite launched to provide commercial ocean color imagery in 1997 (GeoEye, 2008).

Commercial high-resolution remote sensing imagery (see *Commercial Remote Sensing*) became a reality when the former USSR began selling 5 m resolution panchromatic (see *Remote Sensing, Physics and Techniques*) imagery in 1987, followed by Russian sales of 2 m resolution panchromatic imagery in 1992 under the Kosmos/SPIN-2 (Space Information-2) program. Both were digitized from returned film. The US industry felt that SPIN-2 could be the harbinger of a lucrative market that might expand internationally, from which they would be excluded because of existing restrictions on imagery resolution (NASA, 2007). As a result of intense lobbying by industry, in 1994, the first licenses were granted allowing commercial US systems with a resolution of 1 m to be built and flown. Still, in 1998, the highest-resolution satellite imagery commercially

available was from SPIN-2 and from India's IRS-1C and -1D (IRS = Indian Remote Sensing) satellites with 5.8 m panchromatic resolution (Jacobsen, 2008). The first commercial high-resolution remote sensing satellite with 1 m resolution, IKONOS, was launched by the US company Space Imaging (now part of GeoEye; GeoEye, 2008) in 1999, followed by others of similar resolution from several countries (India, Israel, Russia, and South Korea). Those in turn were followed in 2007 by high-resolution SAR systems from Canada (RADARSAT-2 with 3 m resolution) and Germany (TerraSAR-X with 1 m resolution). The commercial high-resolution satellites will be discussed further below. Much of their success has relied on the existence of governments as "anchor-tenants," i.e., principal customers. Sales to the commercial sector have not proven to be as viable as many practitioners envisioned in the 1990s.

Beginning in 1991, remote sensing satellites in the microsatellite (microsat) category (10–100 kg) were launched, followed in 1997 by satellites in the minisatellite (minisat) category (100–500 kg). The first such minisat was NASA's TOMS/Earth Probe carrying the Total Ozone Mapping Spectrometer (TOMS/EP, 2008). These microsats and minisats have enabled much of the international proliferation in this field. Many of the newer countries involved in this field have used the technological know-how of other countries to build much of the space-based hardware. They have used foreign-partnership programs for technology transfer to build up an indigenous capability for constructing the hardware, usually to complement their existing capability in analyzing remote sensing data. A leader in the technology transfer field for nearly 20 years has been the very successful British company SSTL (SSTL, 2008). They have provided microsats to countries including Algeria (Alsat-1/DMC), Chile (FASat-Alfa and Bravo), China (Beijing-1/DMC), Nigeria (NigeriaSat-1/DMC), Portugal (PoSat-1), South Korea (KITSAT-1 and -2), Thailand (TMSat), and Turkey (BilSat-1/DMC). DMC stands for Disaster Monitoring Constellation, an effort in international collaboration established by SSTL for the monitoring of disasters from space using simple multispectral cameras. The former TRW in the USA provided minisats to South Korea (for KOMPSAT-1; Korea Multi-Purpose Satellite-1; KOMPSAT, 2008) and Taiwan (for ROCSAT-1/Formosat-1; Republic of China Satellite-1), used principally for ocean color monitoring (ROCSAT, 2008).

Coupled with the trend toward more small satellites, especially in the minisat class, has been a desire for downsized but highly capable remote sensing instrumentation. Smaller, lighter instruments mean smaller, lighter spacecraft to support them, which means smaller launch vehicles, all equating to lower cost (see *Mission Costs of Earth-Observing Satellites*).

The international proliferation of small satellites has certainly not meant that large satellites have gone out of fashion. In 2002 alone, three high-mass satellites

(eoPortal, 2008) were launched [SPOT-5 from France at 3,000 kg, ADEOS-2 (Advanced Earth Observing Satellite-2) from Japan at 3,700 kg, and Envisat from ESA at 8,000 kg]. Many more medium-to-large (>500 kg) satellites have been launched and are planned.

While many systems hosting multispectral instruments have been launched and are operating successfully, relatively few hyperspectral instruments have been launched. Following some failures and cancelations, the first success was NASA's NMP (New Millennium Program) EO-1 (Earth Observer-1) satellite in 2000, carrying the Hyperspectral instrument with 220 spectral bands (EO-1, 2008). This was followed by the PROBA (Project for On-Board Autonomy) technology demonstration satellite from the UK and Belgium in 2001, carrying the CHRIS (Compact High-Resolution Imaging Spectrometer) hyperspectral imager (CHRIS, 2008). CHRIS data have received broad application. In 2004, NASA's Earth Observing System (EOS) Aura satellite hosted the hyperspectral Ozone Monitoring Instrument (OMI) from the Netherlands and Finland (Aura, 2008). It observes ozone in the ultraviolet (UV) from 0.27 to 0.38 μm (see *Remote Sensing, Physics and Techniques*). Future hyperspectral instruments have been announced by Argentina/Brazil (HSI, 2007), Germany (EnMAP, 2008), India (IMS-1, 2008), Italy (Hypseo, 2007), and South Africa (ZASat-2, 2008). Hyperspectral imagers can quickly generate gigabytes of data and provide a particular challenge for satellite data handling and transmission systems (see *Mission Operations, Science Applications/Requirements and Mission Costs of Earth-Observing Satellites*).

The application of lidar (see *Remote Sensing, Physics and Techniques*) to spaceborne remote sensing remains a fledgling field. The first lidar to fly in space designed to perform remote sensing of the Earth's atmosphere was NASA's LITE (Lidar In-space Technology Experiment) that flew on the shuttle in 1994 (Kramer, 2002). The Russian Balkan-1 lidar was launched on the Mir-Spektr module in 1995 (Kramer, 2002), while the French ALISSA lidar was launched on the Mir-Priroda module in 1996 (Kramer, 2002) to study clouds, aerosols, and the Earth's surface (ALISSA = l'Atmosphere par Lidar Sur SAlout). Currently operating lidars are on the US/French CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation) satellite, launched in 2006, with 100 m laser spot size at the Earth's surface and NASA's EOS ICESat (Ice, Cloud, and land Elevation Satellite), launched in 2003, with 70 m spot size. Planned spaceborne lidars are ESA's in 2008, to measure the tropospheric wind profile for the first time; the European/Japanese EarthCARE (Earth Clouds, Aerosols and Radiation Explorer) satellite, to fly in 2013, with <100 m spot size; and NASA's ICESat-2, to fly in 2015 (Table 1).

Observational platforms

Remote sensing of the Earth may be carried out from towers, balloons, uninhabited airborne vehicles (UAVs),

Observational Systems, Satellite, Table 1 Lidar instruments flying as of 2008 or planned by 2015

Technique	Satellite	Country/organization	Launch date
Backscatter	ICESat	USA	2003
	CALIPSO	USA/France	2006
	EarthCARE	ESA/Japan	2013
Doppler	ADM	ESA	2008

manned aircraft, satellites, and space stations. Because it is impractical to instrument the entire globe with ocean buoys, ground-based sensors, and weather balloons, remote sensing from platforms that can efficiently cover large areas is employed. This is especially true for large remote areas such as the South Pacific. For applications requiring global coverage or large-area synoptic coverage, remote sensing platforms tend to be spaceborne (see *Observational Systems, Satellite and Observational Platforms, Aircraft, and UAVs*). It is physically and financially impractical to provide global coverage with aircraft due to their limited range, limited geographical coverage, weather restrictions, etc. (see *Observational Platforms, Aircraft, and UAVs*).

Platforms that Earth remote sensing has been carried out from include the space shuttle (see *Observational Platforms, Aircraft, and UAVs*). It provided the platform (Kramer, 2002) for the SIR-A, SIR-B, and SIR-C/X-SAR series of SAR instruments (SIR = Shuttle Imaging Radar), as well as the Shuttle Radar Topography Mission (SRTM), a pioneering US/German/Italian Interferometric SAR system (see *Microwave Radiometers, Interferometers*). It was also the platform for LITE, which was used to study clouds and aerosols with lidar (see *Lidar Systems*). The US Skylab space station and the Russian Mir space station both provided platforms for pioneering Earth remote sensing instruments (see *Observational Platforms, Aircraft, and UAVs*). Skylab hosted a microwave imager, scatterometer, and altimeter, among other instruments. The Mir contained approximately 30 remote sensing instruments, housed in five different modules (Kramer, 2002). The International Space Station (ISS) now has the potential to serve as a platform for Earth observation and instrument technology demonstrations.

Satellite orbits: coverage and sampling

Most Earth remote sensing satellites are placed into circular, near-polar, sun-synchronous orbits at altitudes of approximately 600–900 km above the Earth's surface. A satellite in a sun-synchronous orbit passes over a given region on the ground at approximately the same local time every day. This means that the solar illumination angle changes relatively slowly for that area from day to day, which simplifies data interpretation. Some sun-synchronous satellites are in exact repeat orbits, meaning that they fly over the same region at exactly the same local time every day. A single sun-synchronous

LEO satellite carrying a low-resolution instrument with a wide swath providing single-pass coverage over a 3,000 km-wide strip, such as a weather imager, can cover the entire globe in 12 h if the instrument can operate both day and night. Three such satellites in orbits that are spaced 60° apart in longitude can cover the entire globe in approximately 4 h.

A large part of the science community requires frequent imaging revisits of the same area on the Earth. Revisit time is defined here as the average elapsed time between imagery of a given location at the Earth's equator (because for polar sun-synchronous orbits the equator represents the most stressing case). Revisit times are measured for a single instrument onboard the satellite, and they are lower (better) in cases where the instrument can gimbal transverse to the direction of travel of the satellite. An instrument that cannot gimbal is the Thematic Mapper onboard the current US Landsat satellites, in which case 16 days are required for the instrument swath width of 185 km to "paint" the Earth and begin to repeat coverage. The French SPOT system, on the other hand, can gimbal, resulting in revisit times of 3 days. Such a capability is becoming much more commonplace.

Stereo imaging instruments, in which a fore-and-aft image (and an image looking directly down at nadir) of the same region is acquired along the direction of satellite motion (along-track) as the satellite flies over that region, were pioneered in the 1990s by Germany. The MOMS-02/D2 (Modular Optoelectronic Multispectral Scanner) flew on the shuttle and the MOMS-02P instrument flew on the Russian Mir-Priroda space station (Kramer, 2002). These instruments paved the way for the higher-resolution systems with stereo capability that are now flying, many of which acquire stereo imagery by gimbaling a single imager to acquire imagery from different look angles. (Examples are pointed out in the text below.) This capability is useful in mapping, road route selection, communications equipment siting, disaster response, 3-D news "flythroughs," visualization of pieces of real estate, etc.

Some satellites in LEO fly in formation (see *Satellites, Constellations and Formation Flying*). The A-Train is a series of five satellites following one another along the same orbit like "beads on a string" (NASA, 2003). Currently, the NASA EOS Aqua satellite is leading the A-Train, followed by CloudSat, CALIPSO, France's PARASOL (Polarization and Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar), and the NASA EOS Aura satellite (see [International Collaboration](#)). (CloudSat uses radar to study clouds and precipitation, CALIPSO uses lidar to study clouds and aerosols, and PARASOL uses polarization to study clouds and aerosols (PARASOL, 2008). Aqua is a multidisciplinary mission that includes the study of atmospheric water, while Aura studies atmospheric chemistry, trace gases, and pollution.) The temporal gap from Aqua to Aura is 8 min. The last satellite to join the A-Train will be NASA's Orbiting Carbon Observatory (OCO), when it

flies in late 2008 (OCO, 2008). The values of formation flying lie in (1) not investing in a single very large satellite, for which a single problem could result in failure of the entire mission; (2) the ability to acquire synergistic measurements of the same part of the Earth's environment over a very short period of time and combine them to result in an improved understanding of the physics of the Earth's environment, weather, and climate; and (3) sharing of mission cost among agencies (see [Mission Costs of Earth-Observing Satellites](#)).

Some satellites fly as part of a constellation, in which satellites in different orbits, typically carrying the same or similar instruments, operate together to collect data. Flying constellations of satellites allow for more frequent revisit time of the same region, thus faster global coverage. One example is the Disaster Monitoring Constellation described above (see section "[Introduction to Satellite Observational Systems](#)"). More information is available (see [Observational Systems, Satellite](#)).

Many remote sensing weather satellites occupy geosynchronous orbits. Positioned at 36,000 km above the equator, their orbital periods of 24 h make them appear to hang over one spot on the Earth. With broad-area imagers on five "GEO birds," it is possible to cover the globe up to approximately 60° latitude, with spatial resolutions of up to approximately 1 km. Such coverage is provided today from the US Geostationary Operational Environmental Satellites (GOES-East and GOES-West; GOES, 2008) series of metsats (meteorological satellites), the European Meteosat Second Generation (MSG) metsats (MSG, 2008), the Indian InSAT-3 series and their METSAT-1/Kalpana-1 (India, 2008), China's Feng Yun-2 (FY-2) series (FY-2, 2008), and Japan's MTSat (MTSat, 2008) series (the follow-on to their GMS series). South Korea plans to launch a GEO metsat called COMS (Communications, Oceans and Meteorology Satellite) in 2009 (COMS, 2008), while Russia plans to launch GOMS/Elektro-L (Geostationary Orbit Meteorological Satellite) in 2010, replacing GOMS/Elektro-1 that lasted from 1994 to 2000 (GOMS, 2008). This is, in fact, a complete list of environmental remote sensing satellites in, and soon planned for, GEO orbits.

No particular use has been made of medium Earth orbits for remote sensing of the Earth's atmosphere/ocean/land/ice system. MEO has been studied as a possibility for part of a future merged LEO/GEO US weather satellite constellation, but plans have not progressed beyond the study phase.

The Molniya orbit is a type of highly elliptical orbit that allows a satellite to pass slowly over one of the Earth's poles (usually the north pole) and observe the arctic regions for approximately 8 h out of 24 h. A constellation of three such suitably spaced orbits can provide 24 h coverage of the arctic regions. Such orbits are used more for Russian communications satellites than for remote sensing. The upcoming Russian/Finnish Arktika mission for studying the arctic environment plans to fly some of its satellites in Molniya orbits (Arktika, 2008). There are

currently no environmental remote sensing satellites in Molniya orbits.

The L1 libration point, a gravitationally stable point between the Earth and the Sun, has been discussed as a possible location for a satellite to study the Earth as a planet for climate research (the sunward-facing hemisphere of the Earth can be covered). To date, the L1 point has been used for solar-observing satellites. NASA's Triana/DSCOVR mission was envisioned as a climate satellite at L1 (Triana, 2008), but it was canceled as Triana, reinvigorated as DSCOVR, and canceled again. One day, a climate-monitoring satellite may well fly at L1.

Spacecraft subsystems

As outlined in the "Definition" section above, the portion of the satellite minus the instruments is usually called the "spacecraft." The spacecraft exists to support the instruments and provides them with the requisite pointing, commands and data storage, power, thermal conditions, and structural support. Pointing is typically the most difficult requirement to meet, especially for instruments that have very high spatial resolution.

Mission operations

The prime directive of mission operations is to ensure delivery of required data to the user communities of a quality that supports conversion into end products that meet the science requirements (see *Mission Operations, Science Applications/Requirements*). That conversion is typically performed outside of the mission operations environment. An important function of mission operations is to schedule the data the satellite collects in response to user community input and uplink (transmit) that schedule to the satellite. Limitations of onboard data storage or data transmission from the satellite to the ground (downlink) limit the amount of data that can be collected in a given time period. Another important function of mission operations is to monitor the health and status of the instruments and the spacecraft and to take appropriate corrective actions when problems arise. For typical environmental remote sensing satellites in sun-synchronous LEO orbits, the ground receiving stations tend to be in the polar regions, because those satellites can be seen from there on essentially every orbit.

Specific satellite observational systems

We will now examine satellite observational systems, organized by the environmental regime to which they contribute. Many satellites are multidisciplinary and will be so noted. There are several hundred past, present, and future systems that could be included here. The remainder of this entry is restricted to satellites that are actively gathering data in 2008, as well as to ones that appear to be reasonably firmly planned for launch through 2015. A representative sampling of the total number of satellites in each environmental discipline is included (Rao

et al., 1990; Glackin and Peltzer, 1999; Kramer, 2002; eoPortal, 2008; and references below).

Satellite observational systems for the atmosphere

Multidisciplinary

Envisat (Environmental Satellite) is a large European satellite from ESA that has been flying since 2002 (Envisat, 2008). It has application to the atmosphere (clouds, aerosols, ozone, greenhouse gases, and trace gases), the ocean (ocean color, sea surface temperature, sea state, and topography), the land surface (vegetation, snow, and topography), and ice (sea and land ice). Except for certain Russian satellites, this is the most massive satellite that has been flown (8,000 kg) for Earth environmental remote sensing.

EOS Terra, formerly known as EOS AM-1, is a large NASA satellite that has been flying since 1999 (Terra, 2008). It is part of a series of satellites flying in formation, like the A-Train described above, but in a different (morning) orbit. This series includes the Argentinean SAC-C (Satellite de Aplicaciones Cientificas-C; SAC-C, 2008) satellite and the US Landsat-7 and EO-1 (Earth Observer-1), which is a pathfinder for the next Landsat. Terra's name indicates that its applications are focused on the Earth's surface, but it is actually multidisciplinary. It has application to the atmosphere (clouds and aerosols, Earth radiation budget, and pollutants) and the ocean (ocean color) as well.

EOS Aqua, a NASA satellite formerly known as EOS-PM-1, was launched in 2002 and currently leads the A-Train (Aqua, 2008). Aqua's name indicates that its applications are focused on water, but it is actually multidisciplinary. It measures the atmospheric temperature and moisture profiles, precipitation, cloud water, and Earth radiation budget. It measures the ocean surface temperature, wind speed, and ice cover. Over land, it measures soil moisture and snow extent/snow water equivalent.

Megha-Tropiques ("Clouds-Tropics") is a joint mission between the Indian space agency ISRO (Indian Space Research Organization) and the French space agency CNES (Centre National d'Etudes Spatiales) (see Megha-Tropiques, 2008). It is a satellite for monitoring of the tropics that is scheduled for launch in 2009. It hosts the ScaRaB (Scanner for Radiation Budget) radiometer and the MADRAS conically scanning microwave radiometer (MADRAS = Microwave Analysis and Detection of Rain and Atmospheric Structures).

NPOESS (National Polar-orbiting Operational Environmental Satellite System) is a joint program between NOAA (National Oceanic and Atmospheric Administration), the US DoD (Department of Defense), and NASA (NPOESS, 2008). EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) is a partner and will provide the European MetOp (Meteorological Operational) satellite (MetOp, 2008) as part of the NPOESS constellation of three satellites. The NPOESS

program was established in 1994. The first satellite is currently scheduled to fly in 2013. Originally a joint NOAA-DoD program focused on operational weather and environmental data, NPOESS will now also host selected climate-monitoring instruments and will serve as a partial follow-on to NASA's EOS program. As presently planned, NPOESS will measure more than 35 different environmental parameters (atmosphere, ocean, land, and ice).

NASA's NPP (NPOESS Preparatory Project) satellite is intended to be a technology demonstrator for NPOESS, as well as a gap-filler between the projected end-of-life of NASA's EOS satellites and NPOESS (NPP, 2008). Launch is scheduled for 2010. NPP will provide a subset of the NPOESS environmental parameters.

Cloud characteristics and aerosols

AIM (Aeronomy of Ice in the Mesosphere) is a NASA satellite that was launched in 2007 to study polar mesospheric clouds, also known as noctilucent ("night shining") clouds (AIM, 2008). Occurring at altitudes around the mesopause (approximately 84 km above ground level), they remain a poorly understood phenomenon.

CloudSat is a NASA satellite that was launched in 2006 and is part of the A-Train (CloudSat, 2008). Its mission is to study the structure of clouds, which is very important in understanding climate change. It carries the first spaceborne cloud profiling radar, operating at 94 GHz, which is a frequency that can penetrate most clouds (see *Cloud properties*).

CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation), a joint mission between NASA and CNES, was launched in 2006 and is part of the A-Train (CALIPSO, 2008). Its mission is to study clouds and aerosols, and its payload includes a backscatter lidar (see *Aerosols*).

EarthCARE (Earth Clouds, Aerosols and Radiation Explorer) is a joint mission between ESA and the Japanese space agency JAXA (Japanese Aerospace Exploration Agency), formerly known as NASDA. Currently scheduled for launch in 2013, its payload will include a cloud radar and a cloud-aerosol lidar (EarthCARE, 2008).

Stratospheric ozone, greenhouse gases, and trace gases
EOS Aura, a NASA satellite formerly known as EOS Chem, was launched in 2004 and is the last satellite in the A-Train (Aura, 2008). It has application to ozone and greenhouse gases, as well as a variety of trace gases and pollutants in the troposphere and stratosphere (see *Trace Gases, Troposphere - Detection from Space and Trace Gases, Stratosphere, and Mesosphere*).

Odin is a joint Swedish/Canadian satellite (Odin, 2008) from the Swedish National Space Board (SNSB) and the Canadian Space Agency (CSA). Launched in 2001, its purpose is to measure ozone and chlorine chemistry. Its payload includes a sub-mm radiometer (see *Remote Sensing, Physics and Techniques*).

SCISAT-1 is a CSA satellite launched in 2003 to study ozone chemistry (SCISAT, 2006). Its payload is a Fourier Transform Spectrometer (see *Remote Sensing, Physics and Techniques*).

OCO (Orbiting Carbon Observatory) is a NASA satellite to be launched in 2008 to study atmospheric carbon dioxide (OCO, 2008). It will be the last satellite to join the A-Train. The payload is comprised of a grating spectrometer (see *Remote Sensing, Physics and Techniques*).

Glory is a NASA satellite to be launched in 2009 to study atmospheric aerosols and black carbon (soot). (See Glory, 2008.) Its payload will consist of an imaging polarimeter (see *Remote Sensing, Physics and Techniques*).

GCOM-A1 (Global Change Observation Mission-A1) is a JAXA satellite to be launched in 2008 to study ozone and greenhouse gases (GCOM, 2008). Its payload will consist of a UV spectrometer and a solar occultation spectrometer (see *Remote Sensing, Physics and Techniques*).

Rainfall

TRMM (Tropical Rainfall Measuring Mission) is a joint JAXA/NASA satellite launched in 1997 (TRMM, 2008) to study tropical rainfall (see *Rainfall*). Its payload includes a precipitation radar.

GPM (Global Precipitation Measurement) mission consists of a "core" JAXA/NASA satellite to be launched in 2013 and a series of "constellation" satellites. The core satellite's payload will include a precipitation radar (GPM, 2008).

Earth radiation budget

No satellites have been dedicated to this application since NASA's ERBS (Earth Radiation Budget Satellite) that was launched in 1984 (Kramer, 2002). These types of measurements are typically carried out by individual instruments onboard multipurpose satellites. The instruments include NASA's CERES (Clouds and the Earth's Radiant Energy System) on Terra, Aqua, and NPOESS, the Russian ScaRaB on Megha-Tropiques, and ESA's GERB (Geostationary Earth Radiation Budget) on the geosynchronous weather satellite MSG (see *Earth Radiation Budget, Top-of-Atmosphere Radiation*).

Lightning

This is also detected using individual instruments onboard multipurpose satellites. NASA's LIS (Lightning Imaging Sensor) is flying on TRMM, and NASA's GLM (GEO Lightning Mapper) will fly on NOAA's future GOES-R (GOES-R, 2008) weather satellites (see *Lightning*).

Tropospheric winds

ESA's ADM (Atmospheric Dynamics Mission; ADM, 2008) will be the first satellite to host a Doppler wind lidar to measure the vertical wind profile in the troposphere (see *Remote Sensing, Physics and Techniques*). Scheduled for launch in 2008, this pioneering mission will host an instrument of a type that has been under study since 1978 (see *Tropospheric Winds*).

GPS occultation and reflectometry

ROCSAT-3/COSMIC (Republic of China Satellite-3/ Constellation Observing System for Meteorology, Ionosphere, and Climate; COSMIC, 2008) is a joint Taiwanese/US constellation of satellites from Taiwan's National Space Organization (NSPO) and the US University Corporation for Atmospheric Research (UCAR). Launched in 2006, this constellation of six microsatellites measures the ionospheric electron density profile, the atmospheric temperature profile (see *Land Surface Temperature*), and other parameters, using the GPS occultation technique (see *Remote Sensing, Physics and Techniques*).

FEDSAT-1 (Federation Satellite-1) is an Australian satellite from the Commonwealth Scientific and Industrial Research Organisation (CSIRO). Launched in 2002, it performs GPS occultation (Fedsat, 2002).

CHAMP (CHALLENGING Minisatellite Payload) is a joint German/US mission (CHAMP, 2008) from the German Space Agency (DARA), the German Aerospace Center (DLR), the German Research Centre for Geosciences (GFZ), and NASA's JPL (Jet Propulsion Laboratory). Launched in 2000, it was the first satellite to include a capability for GPS reflectometry as well as GPS occultation (see *Remote Sensing, Physics and Techniques*).

SAC-C is a joint mission between Argentina's National Commission for Space Activities (CONAE) and NASA (SAC-C, 2008). Launched in 2000, it was the second satellite to include a capability for GPS reflectometry as well as GPS occultation.

Weather prediction and severe storms

LEO satellites for weather prediction and a host of other environmental measurements (see *Weather Prediction*) include NPP and NPOESS, described above (see section on "Multidisciplinary"). The two programs that were originally merged to create NPOESS are the DMSP Block 6 (Defense Meteorological Satellite Program) system (not discussed here) and NOAA's POES O-P-Q (Polar-orbiting Operational Environmental Satellite). POES (POES, 2008) has application to the atmosphere, oceans, land, and ice. POES/NOAA-18 was launched in 2005, and the last satellite in the series (dubbed NOAA-N-prime before launch) is scheduled for launch in 2009. The ESA/EUMETSAT MetOp satellite program was described above. It is an international collaboration with NPOESS and POES. MetOp-1 was launched in 2006 into a midmorning orbit to complete the present POES constellation (IJP, 2008). MetOp-2 and MetOp-3 are planned for launch in 2010 and 2014, during the transition from POES and DMSP to NPOESS.

Russia has flown the Meteor series of LEO metsats since 1969, the most recent being Meteor-3 M-1 from Roskosmos/Roshydromet, launched in 2001 (Russia, 2003). Meteor-M-N1, the first of a new series, is scheduled for launch in 2009. China has flown the FY-1 (Feng Yun-1) series of LEO metsats since 1988, the most recent

being in 2002 (FY, 2008). FY-3A, the first of a new series, was launched in 2008.

Present and near-future (through 2010) GEO weather satellites for monitoring weather and severe storms (see *Severe Storms*) are enumerated above in the section "Satellite Orbits: Coverage and Sampling." Regarding the next generation, ESA and EUMETSAT plan to launch Meteosat Third Generation (MTG) satellites starting in 2015, China plans to launch Feng Yun-4 (FY-4) satellites starting possibly in 2012 (FY-4, 2008), and the USA plans to launch GOES-R in 2015 (GOES-R, 2008).

Satellite observational systems for the oceans

Multidisciplinary

Aqua, Envisat, NPOESS, NPP, and Terra are described above. Please see the section on "Multidisciplinary" under the section "Satellite Observational Systems for the Atmosphere."

ESA's ERS-2 (European Remote Sensing-2) is a large oceanographic satellite launched in 1995 (ERS, 2008). One instrument combines the functions of a SAR and a scatterometer (see *Remote Sensing, Physics and Techniques*). The rest of the payload includes a radar altimeter, a sea surface temperature imager, and an ozone monitor.

ESA's Sentinel-1 satellite is part of Europe's Global Monitoring for Environment and Security (GMES). Scheduled for launch in 2011, it will host a SAR, as a follow-on to ERS and Envisat (Sentinel, 2008). ESA's Sentinel-3 satellite is scheduled for launch in 2012. It will measure ocean color and surface temperature as well as land vegetation and surface temperature.

China's HY-2 (Hai Yang-2) from the Chinese Association for Science and Technology (CAST) is due to be launched in 2009 (HY, 2007). It will host two ocean color imagers, a microwave radiometer, an altimeter, and a scatterometer.

Table 2 summarizes passive microwave imagers, as well as polarimetric radiometers (see next section).

Sea surface wind vector

Historically, instruments that measure both sea surface wind speed and direction have been active scatterometers. More recently, a proof-of-concept mission has been flown to examine the viability of making this measurement with a passive polarimetric microwave radiometer (see *Reflected Solar Radiation Sensors, Polarimetric*). As of 2008, it appears that each technique has advantages, depending on the range of wind speeds.

ESA's ERS-2 AMI (Active Microwave Instrument) can operate as a SAR or a scatterometer, as noted above (ERS, 2008). It has been providing sea surface wind data part-time since 1995. The ESA/EUMETSAT MetOp-1 satellite hosts the ASCAT (Advanced Scatterometer), which is a follow-on to the AMI (see *Sea Surface Wind/Stress Vector*).

NASA's QuikSCAT satellite, launched in 1999, hosts the SeaWinds scatterometer (QuikSCAT, 2008). Its data

Observational Systems, Satellite, Table 2 Passive microwave imagers and wind polarimeters flying as of 2008 or planned by 2015

Technique	Satellite	Country/ organization	Launch date
Imager	Aqua	USA	2002
	FY-3A	China	2008
	Oceansat-1	India	1999
	TRMM	Japan/USA	1997
	Aquarius	USA/Argentina	2010
	FY-3B, C, ...	China	2010+
	GCOM-B1	Japan	2008
	GPM Core	Japan/USA	2013
	HY-2	China	2009
	Jason-2	USA/France/ EUMETSAT	2008
	Megha-Tropiques	India/France	2009
	NPOESS-C2	USA	2016
	Oceansat-2	India	2008
Wind polarimeter	SMAP	USA	2012
	Coriolis/ WindSat	USA	2003
	NPOESS-C2	USA	2016
Interferometer	SMOS	ESA	2009

have proven to be useful for a variety of oceanographic applications.

The Coriolis satellite from the US Naval Research Laboratory (NRL) was launched in 2003 to test the microwave passive polarimetric approach from space (Coriolis, 2008). The WindSat instrument on Coriolis continues to provide wind speed and direction data, as well as other environmental parameters. Coriolis is a risk reduction mission for the NPOESS MIS (Microwave Imager/Sounder), which may include a passive polarimetric capability for wind.

India's IRS-2B/Oceansat-2 from ISRO is due for launch in 2008 (India, 2008). It will carry a "pencil-beam" scatterometer (see *Radar, Scatterometers*) and an ocean color imager.

Other future scatterometers are planned by Argentina, China, and Japan.

Ocean color

OrbView-2, launched in 1997 by the US company ORBIMAGE (now GeoEye), has provided ocean color imagery from its SeaWiFS (Sea-viewing Wide-Field Sensor) instrument (OrbView-2, 2008). This imagery provides a measure of organic material (such as phytoplankton) and marine optical properties (see *Ocean Measurements and Applications, Ocean Color*).

ROCSAT-1/Formosat-1 (Republic of China Satellite-1) from Taiwan's NSPO was launched in 1999 (ROCSAT, 2008). This satellite, Taiwan's first remote sensing satellite, is dedicated to ocean color imagery.

China's HY-1B from CAST was launched in 2007 (HY, 2007). Its payload is dedicated to ocean color imagery. The follow-on will be the HY-2 series, scheduled for first launch in 2009.

India's IRS-P4/Oceansat-1 from ISRO was launched in 1999 with an ocean color imager (Oceansat, 2008). The follow-on Oceansat-2 will also host an ocean color imager.

Sea surface temperature

This parameter (see *Sea Surface Temperature*) is typically measured by instruments flying on multidisciplinary satellites. Examples, all described above, are Aqua, Envisat, ERS-2, MetOp, NPOESS, NPP, and Sentinel-3.

Sea surface topography

Jason-1/TPFO (TOPEX/Poseidon Follow-On), a joint mission between CNES, NASA, and NOAA, was launched in 2001 (Jason, 2008). It hosts an altimeter for measurement of sea surface topography (see *Ocean Surface Topography*) and a microwave radiometer for the measurement and removal of atmospheric water vapor effects.

Jason-2/OSTM (Ocean Surface Topography Mission) is scheduled to be launched in 2008 (Jason, 2008). It will carry the same payload and adds EUMETSAT as a partner.

Table 3 summarizes active microwave instruments.

Sea surface salinity

ESA's SMOS (Soil Moisture/Ocean Salinity) will be the first satellite to measure ocean salinity (see *Sea Surface Salinity*). It will also be the first interferometric microwave radiometer (using aperture synthesis) in space (see *Remote Sensing, Physics and Techniques*). Operating at 1.4 GHz, it will measure not only ocean salinity but also soil moisture, which is best measured at this relatively low frequency. SMOS is scheduled for launch in 2009 (SMOS, 2008).

Aquarius is a joint mission between the USA and Argentina (NASA and CONAE). It will host a large real-aperture 1.4 GHz microwave radiometer for the measurement of ocean salinity and is scheduled for launch in 2010 (Aquarius, 2008).

Satellite observational systems for the land surface

Aqua, Envisat, NPOESS, NPP, and Terra are described above. Please see the section on "Multidisciplinary" under the section "Satellite Observational Systems for the Atmosphere."

Earth resources and environmental monitoring

Landsat is the longest-lived series of environmental remote sensing satellites (Landsat, 2008). These and several others listed in this section fill a niche between the low-resolution land imagers on satellites, such as Envisat, Terra, Aqua, NPOESS, and MetOp, and the high-resolution land imagers on commercial high-resolution satellites. Landsat-5 and Landsat-7 from NASA and the

Observational Systems, Satellite, Table 3 Active microwave instruments flying as of 2008 or planned by 2015

Technique	Satellite	Country/ organization	Launch date
Altimeter	Envisat	ESA	2002
	ERS-2	ESA	1995
	Geosat Follow-On	USA	1998
	Jason-1	France/USA	2001
	Cryosat-2	UK	2009
	HY-2	China	2009
	Jason-2	USA/France/ EUMETSAT	2008
Scatterometer	Sentinel-3	ESA, European Commission	2012
	ERS-2	ESA	1995
	MetOp-1	ESA, EUMETSAT	2006
	QuikSCAT	USA	1999
	Aquarius	USA/Argentina	2010
	GCOM-B1	Japan	2008
	HY-2	China	2009
	Oceansat-2	India	2008
	CloudSat	USA	2006
	TRMM	Japan/USA	1997
Real-aperture Radar	EarthCARE	ESA/Japan	2013
	GPM Core	Japan/USA	2013
	SMAP	USA	2012
	ALOS	Japan	2006
SAR	Cosmo/SkyMed-1, 2	Italy	2007
	Envisat	ESA	2002
	ERS-2	ESA	1995
	RADARSAT-1	Canada	1995
	RADARSAT-2	Canada	2007
	Remote Sensing-1	China	2006
	Remote Sensing-2,3	China	2007
	TerraSAR-X	Germany	2007
	Cosmo/SkyMed-3,4	Italy	2008
	HJ-1C	China	2009
	HJ-2	China	2010
	KOMPSAT-5	South Korea	2008
	MASAR	Brazil/Germany	2013
	Paz	Spain	2012
	RADARSAT-3	Canada	2013
	RISAT	India	2008
	SAOCOM-1A	Argentina/Italy	2010
	Sentinel-1	ESA, European Commission	2011
	TanDEM-X	Germany	2009
InSAR	TerraSAR-X with TanDEM-X	Germany	2009
GPS-based	CHAMP	Germany/USA	2000
	Fedsat-1	Australia	2002
	MetOp-1	ESA, EUMETSAT	2006
	ROCSAT-3/ COSMIC	Taiwan/USA	2006
	SAC-C	Argentina/USA	2000
	UK-DMC	UK	2003
	Oceansat-2	India	2008

USGS (United States Geological Survey) are medium-resolution Earth resources systems launched in 1984 and 1999 and still operating in 2008. The Enhanced Thematic Mapper-plus (ETM+) on Landsat-7 provides 15 m panchromatic, 30 m multispectral, and 60 m LWIR (long-wave infrared) imagery over a 185 km-wide swath. The Landsat Data Continuity Mission (LDCM), to be launched in 2011, is a follow-on with a more advanced instrument, the Operational Land Imager (LDCM, 2008).

Russia began flying the Resurs-O1 (Resurs = "Resource") series of Earth resources satellites in 1985 (Resurs 2008). It appears that none remain operational (Resurs History, 2008).

The French SPOT series of satellites has been flying since 1986. Up until SPOT-5, they provided 10 m panchromatic and 20 m multispectral imagery. SPOT-5 provides 5 m panchromatic imagery that can be double-sampled and processed to approximate about 3 m resolution imagery (Jacobsen, 2008). The instrument can acquire stereo imagery and can gimbal, unlike Landsat. The follow-on SPOT-6 is not yet firmly scheduled.

The Indian IRS series of Earth resources satellites from ISRO has been flying since 1988. IRS-1D, launched in 1997, has 5.8 m panchromatic, 23.5 m multispectral (vis-near IR), and 70.5 m short-wave IR resolution (India, 2008). Resourcesat-1, the first follow-on mission launched in 2003, has 5.8 m multispectral (vis-near IR) resolution among others, with gimbaling and stereo capability (Resourcesat, 2008). Resourcesat-2 is scheduled for launch in 2009.

Japan has been flying Earth resources satellites since 1996. Still operating in 2008 is ALOS (Advanced Land Observing Satellite) that was launched in 2006 (ALOS, 2008). It can acquire 2.5 m stereo imagery with one instrument and 10 m multispectral imagery with another instrument, which can also gimbal. GCOM-B1, to be launched in 2008, is the follow-on satellite (GCOM, 2008).

China and Brazil have jointly been flying Earth resources satellites since 1999, the CBERS/ZY1 series (CBERS = China-Brazil Earth Resources Satellite; ZY = Zi Yuan), from CAST and the Brazilian National Institute of Space Research (INPE). CBERS-2B, launched in 2007, has 2.7 m panchromatic and 20 m multispectral instruments (CBERS, 2008). The follow-on satellite, CBERS-3, is scheduled to launch in 2008 and will have 5 m panchromatic and 20 m multispectral capability. CBERS-4, -5, and -6 are planned for launch after 2010.

Six other nations plan to launch Earth resources satellites soon, five of them in 2008. X-Sat, from the Nanyang Technological University (NTU) in Singapore, will provide 10 m multispectral resolution (X-Sat, 2003). ZASat-1/Sumbandila, from the University of Stellenbosch in South Africa, will provide 6.5 m multispectral resolution (ZA = Zuid Afrika; Sumbandila = "Lead the Way"). Its follow-on, ZASat-2, planned for 2009, will provide 2.5 m panchromatic, 5 m multispectral, and 15 m hyperspectral imagery (ZASat, 2008). THEOS (Thailand Earth Observation System) from Thailand's GISTDA

(Geo-Informatics and Space Technology Development Agency) will provide 2 m panchromatic and 15 m multispectral imagery (THEOS, 2008). Razaksat, from ATSB (Astronautic Technology Sdn Bhd) in Malaysia, will provide 2.5 m panchromatic and 5 m multispectral imagery (Razaksat, 2005). RASAT, from TUBITAK (The Scientific and Technological Research Council of Turkey) in Turkey, will provide 7.5 m panchromatic and 15 m multispectral imagery (RASAT, 2006). Finally, the first of the SAOCOM (Satellites for Observation and Communication) series of satellites from CONAE in Argentina (SAOCOM-1A) is scheduled for launch in 2010, carrying a SAR with 7–100 m resolution (SAOCOM, 2008).

Disaster management

SSTL's Disaster Monitoring Constellation (DMC) satellite series was described above. (See section "[Introduction to Satellite Observational Systems](#).")

India's *RISAT* (Radar Imaging Satellite) is due to launch in 2008. It will carry a SAR with 3 to 50 m resolution (RISAT, 2008). Its stated applications are disaster monitoring and crop inventory.

China's *Remote Sensing Satellite* (YaoGan WeiXing) series consists of a constellation of three reportedly disaster-monitoring satellites hosting SARs with unknown spatial resolution. These were launched in 2006–2007 (YaoGan, 2008).

China's CEMD/Huanjing (*HJ-1* series) consists of a constellation of three reportedly disaster-monitoring satellites, two carrying multispectral imagers (HJ-1A and -1B) and one carrying a SAR (HJ-1C). HJ-1A is scheduled for launch in 2008 (CEMD, 2008).

Soil moisture and freeze/thaw state

ESA's *SMOS* was described above. See the section "[Sea Surface Salinity](#)."

NASA's *SMAP* (Soil Moisture Active/Passive) satellite will exploit the synergy between active and passive microwave remote sensing techniques for measuring soil moisture (see [Soil Moisture](#)) and surface freeze/thaw state by carrying a 1.4 GHz radar and a large real-aperture 1.4 GHz microwave radiometer (SMAP, 2008). SMAP is scheduled for launch in 2012.

Commercial and hybrid government/commercial high-resolution systems, electrooptical

The genesis of the commercial high-resolution remote sensing industry is discussed above (see section "[Introduction to Satellite Observational Systems](#)"). Today, these systems typically have the ability to gimbal to provide frequent imagery of a given region, the ability to cover large areas efficiently, and the ability to collect stereo imagery, which can be converted to topographic information. Potential commercial applications (see [Commercial Remote Sensing](#)) include mapping, precision farming, urban planning, forestry, oil and mineral exploration, emergency response, disaster assessment, route selection for roadways, communications siting, real estate, land

use, utility pipeline and long-distance power line monitoring, news reporting, commodities trading, and many aspects of the insurance industry.

We present these satellites in chronological order of successful launch. Note that most, if not all, of these systems may be classified as "dual-use" (commercial/military), since governmental military organizations typically act as anchor-tenants (primary customers) for these systems (Jacobsen, 2008).

IKONOS-1 from the US firm Space Imaging (now GeoEye; GeoEye, 2008) was launched in 1999 (Ikonos is Greek for "Image"). With 0.8 m panchromatic and 4 m multispectral resolution, it is still operating. *EROS-A* from the Israeli firm ImageSat International (EROS, 2008), with 1.8 m panchromatic imagery, was launched in 2000, while *EROS-B* with 0.7 m panchromatic imagery was launched in 2006 (EROS = Earth Resources Observation Satellite). *QuickBird* from the US firm Digital Globe, with 0.6 m panchromatic and 2.5 m multispectral imagery, was launched in 2001 (Quickbird, 2008). *Cartosat-1/IRS-P5* from ISRO, with 2.5 m panchromatic resolution, was launched in 2005 (Cartosat, 2008), followed by *Cartosat-2/IRS-2A* with 1 m resolution in 2005 (Cartosat-2, 2008). [Note that commercialization of Cartosat imagery may not come about (GeoInformatics, 2007)]. *Resurs-DK1*, from the Russian organization Rosaviakosmos, with 1 m panchromatic and 3 m multispectral resolution (Russia, 2003), and perhaps the first Russian digital downlink system (replacing film canisters), was launched in 2006. *KOMPSAT-2/Arirang-2*, from the Korea Aerospace Research Institute (KARI), with 1 m panchromatic and 4 m multispectral resolution, was also launched in 2006 (KOMPSAT, 2008). ("Arirang" is the most popular Korean folk song.) *ROCSAT-2/Formosat-2* from the NSPO in Taiwan, with 2 m panchromatic and 8 m multispectral resolution, was launched in 2004 (ROCSAT, 2008). *KOMPSAT-2* and *ROCSAT-2* images are available commercially from SPOT Image. Finally, *WorldView-1*, from the US firm Digital Globe, with 0.5 m panchromatic resolution, was launched in 2007 (Digital Globe, 2008).

Future missions include *GeoEye-1* with 0.4 m panchromatic resolution in 2008 and *GeoEye-2* with 0.25 m panchromatic resolution in 2011, from the US firm GeoEye (GeoEye, 2008). *WorldView-2* with 0.5 m panchromatic resolution is scheduled in 2008 (Digital Globe/WV-2, 2008). *EROS-C* with 0.7 m panchromatic resolution is scheduled for 2008. (All of these systems also have multispectral capability.) *Cartosat-3* may have 0.30 m panchromatic resolution and was launched in 2011 (Cartosat-3, 2008). *SPOT-6* from CNES, with 2.5 m panchromatic resolution, is not yet firmly scheduled.

These systems are summarized in [Table 4](#).

Commercial and hybrid government/commercial high-resolution systems, SAR

TerraSAR-X is a German SAR satellite from the DLR and the firm EADS Astrium GmbH (EADS = European

Observational Systems, Satellite, Table 4 High-resolution commercial systems flying as of 2008 or planned by 2015

Technique	Satellite	Country/ organization	Launch date
Electrooptical	Cartosat-1	India	2005
	Cartosat-2	India	2005
	Cartosat-3	India	2011
	EROS-A	Israel	2000
	EROS-B	Israel	2006
	IKONOS-1	USA	1999
	KOMPSAT-2	South Korea	2006
	Quickbird	USA	2001
	ROCSAT-2	Taiwan	2004
	Resurs-DK1	Russia	2006
	WorldView-1	US	2007
	EROS-C	Israel	2008
	GeoEye-1	USA	2008
	GeoEye-2	USA	2011
	SPOT-6	France	?
	WorldView-2	USA	2008
Radar (SAR)	RADARSAT-2	Canada	2007
	TerraSAR-X	Germany	2007
	TanDEM-X	Germany	2009

Aeronautic and Space Company). Launched in 2007, it has a spatial resolution of 1 m (TerraSAR-X, 2008). It is designed to fly in formation with the German SAR system TanDEM-X, from the same organizations, with an identical SAR instrument (TanDEM-X, 2008). Potential commercial applications include ship routing through sea ice and oil spill assessment (see *Commercial Remote Sensing*).

RADARSAT-2 is a Canadian SAR satellite from CSA and MDA (MacDonald, Dettwiler and Associates, Ltd.) with 3 m resolution and a capability to observe with horizontal and vertical polarization in all send-receive combinations (RADARSAT, 2008). This is a follow-on to the lower-resolution single-polarization RADARSAT-1 and was launched in 2007. RADARSAT-3 is still in the formulation stages.

Other high-resolution systems

There are other high-resolution electrooptical and SAR systems, some of which have been announced as dual-use (civil/military) systems.

The Italian COSMO/SkyMed (Constellation of small Satellites for Mediterranean basin Observation) hosts a 1 m resolution X-band SAR and has been announced as a dual-use system, to operate in conjunction with the French Pleiades system. The first two satellites of the planned four-satellite constellation were launched in 2007 (COSMO/SkyMed, 2008). The French Pleiades satellite hosts an instrument with a 0.7 m resolution panchromatic band and 2 m resolution multispectral capability (Pleiades, 2008). It has also been announced as a dual-use

system. Two satellites are scheduled for launch in 2008 and 2010.

The Spanish Paz satellite will host a high-resolution SAR, to be launched in 2012. The Spanish Ingenio satellite will host a 2.5 m panchromatic imager, also to be launched in 2012. Both systems have been announced as dual-use (Paz/Ingenio, 2008).

Pakistan is planning to launch two RSSS (Remote Sensing Satellite System) satellites with 4 m panchromatic and 10 m multispectral capability (RSSS, 2008). This has been announced as a dual-use system.

Egypt launched the high-resolution MisrSat-1 ("EgyptSat-1") in 2007, but few details are available (MisrSat, 2008). China launched a series of three satellites called China Resource-1/ZY2-1 through China Resource-3/ZY2-3 in 2000–2004 with 3 m panchromatic imaging, but few details are available (China Resource, 2008). (This series should not be confused with the CBERS/ZY1 series.)

Other satellites that may have potential as dual-use systems include India's Cartosat and RISAT, Malaysia's Razaksat, South Africa's ZASat-2, and Thailand's THEOS (all are described above).

Commercial agriculture

The RapidEye constellation of five satellites from the company RapidEye AG in Germany is scheduled to be launched in 2008 (RapidEye, 2008). Its instruments will produce 6 m resolution multispectral imagery for application to precision farming and crops (see *Precision Agriculture and Crop Stress*).

Hyperspectral systems

Hyperspectral systems are summarized above in the section "[Introduction to Satellite Observational Systems](#)."

Satellite observational systems for the solid Earth

These two satellites are examples of ones that are capable of addressing applications discussed in this volume (see also the volume *Solid Earth Geophysics, Measurements and Applications* and its sections *Geodesy, Crustal Deformation, Subsidence, Volcanism and Landslides*). The German TerraSAR-X and TanDEM-X satellites are described above (see section "[Commercial and Hybrid Government/Commercial High-Resolution Systems, SAR](#)"). Flying in formation, they will perform single-pass Interferometric SAR (InSAR) and will reportedly be able to map the entire Earth in 2.5 years.

Gravity field and water resources

The GRACE (Gravity Recovery and Climate Experiment) mission from NASA and the DLR was launched in 2002 (GRACE, 2008). Two satellites are flying in formation, and their detailed trajectories in space can be used to measure changes in the Earth's gravitational field (see *Gravity Field*). This in turn can be used to infer changes in the mass of arctic ice and in the distribution of fresh water (see *Water Resources*). ESA's GOCE (Gravity field and

steady-state Ocean Circulation Explorer) mission is comprised of a single satellite (GOCE, 2008). It is scheduled for launch in 2008.

Satellite observational systems for the cryosphere (Ice)

The following satellites address applications discussed in this volume (see *Cryosphere, Measurements and Applications* and its sections *Ice Sheets*, *Sea Ice Extent*, and *Concentration*, *Polar Ice Dynamics*, and *Polar Ocean Navigation*). ESA's Envisat is described above. Please see the section on "Multidisciplinary" under the section "Satellite Observational Systems for the Atmosphere" in this entry. ESA's Sentinel-1 is described above; see the section "Multidisciplinary" under the section "Satellite Observational Systems for the Ocean" in this entry.

NASA's EOS ICESat, discussed above (see section "Introduction to Satellite Observational Systems"), hosts a backscatter lidar for measuring ice, land, and cloud elevation (ICESat, 2008). The follow-on mission IceSat-2 is scheduled for launch in 2015. Canada's RADARSAT-1 and RADARSAT-2 are discussed above (see section "Commercial and Hybrid Government/Commercial High-Resolution Systems, SAR"). Their SAR systems are good for providing ice data day and night. The Russian/Finnish Arktika satellite constellation, scheduled for launch starting in 2010, is designed to provide full-time coverage of the Arctic (Arktika, 2008). CryoSat-2 from the UK, scheduled for launch in 2009, will carry an interferometric altimeter to determine the thickness of land and sea ice (CryoSat-2, 2008).

Conclusion

The field of remote sensing of the Earth's environment from space is one that continues to expand internationally and that has benefited from advancements in instrument, spacecraft, ground-based data handling, and conversion of data and images to information. The availability of trained scientists and engineers who can understand the data and bring an end-to-end systems perspective to the field is a continuing issue. Training in astronomy and astrophysics provides a nearly ideal background for working in Earth remote sensing (Wertz, 1986). It has become a very common career path and a natural transition from a field with relatively limited job opportunities to one with excellent career potential and intellectual challenge.

Bibliography

- ADM, 2008. http://www.esa.int/esaLP/ESAFN52VMOC_LPadmaecolus_1.html.
 AIM, 2008. <http://aim.hamptonu.edu/>.
 ALOS, 2008. http://www.jaxa.jp/projects/sat/alos/index_e.html.
 Aqua, 2008. <http://aqua.nasa.gov/>.
 Aquarius, 2008. <http://aquarius.gsfc.nasa.gov/>.
 Arktika, 2008. http://www.spacedaily.com/reports/Russia_To_Launch_Space_Project_To_Monitor_The_Arctic_In_2010_999.html.

- Aura, 2008. <http://aura.gsfc.nasa.gov/>.
 CALIPSO, 2008. <http://www-calipso.larc.nasa.gov/>.
 Cartosat, 2008. <http://www.isro.org/Cartosat/Page3.htm>.
 Cartosat-2, 2008. <http://www.isro.gov.in/pslv-c7/pg6.html>.
 Cartosat-3, 2008. <http://www.thehindu.com/2008/04/28/stories/2008042851471000.htm>.
 CBERS, 2008. http://www.cbbers.inpe.br/en/index_en.htm.
 CEMD, 2008. http://www.spacemart.com/reports/China_To_Launch_1st_Environment_Monitoring_Satellite_999.html.
 CHAMP, 2008. <http://op.gfz-potsdam.de/champ/>.
 China Resource, 2008. <http://www.cira.colostate.edu/cira/RAMM/hillger/environmental.htm#crs>.
 CHRIS, 2008. http://www.esa.int/esaEO/SEMLFM2VQUD_index_0_m.html.
 CloudSat, 2008. <http://CloudSat.atmos.colostate.edu/>.
 COMS, 2008. <http://rsd.gsfc.nasa.gov/goes/text/geonews.html#COMS>.
 Coriolis, 2008. <http://www.nrl.navy.mil/content.php?P=04REVIEW87>.
 COSMIC, 2008. <http://www.cosmic.ucar.edu/>.
 COSMO/SkyMed, 2008. <http://www.asi.it/SiteEN/ContentSite.aspx?Area=Osservare%20la%20Terra>.
 Cryosat-2, 2008. <http://www.esa.int/esaLP/LPcryosat.html>.
 Digital Globe, 2008. <http://www.digitalglobe.com/>.
 Digital Globe/WV-2, 2008. <http://www.digitalglobe.com/index.php/88/WorldView-2>.
 EarthCARE, 2008. http://www.esa.int/esaCP/SEMMBQ1YUFF_index_0.html.
 EnMAP, 2008. <http://www.enmap.org>.
 Envisat, 2008. <http://envisat.esa.int/>.
 EO-1, 2008. <http://eo1.gsfc.nasa.gov/>.
 eoPortal, European Space Agency, 2008. <http://directory.eoportal.org>.
 EROS, 2008. <http://www.imagesatintl.com/default.asp?catid=%7BFD649417-4866-45CE-96F1-139BD1409DCE%7D>.
 ERS, 2008. <http://earth.esa.int/ers/>.
 Fedsat, 2002. www.skyrocket.de/space/doc_sdat/fedsat-1.htm.
 FY, 2008. http://209.85.141.104/search?q=cache:aQvDvPpK2oYJ:www.wmo.ch/pages/prog/gcos/aopXIV/22_CMA_SatProducts.pdf+FY-1+Satellite&hl=en&ct=clnk&cd=11&gl=us.
 FY-2, 2008. <http://rsd.gsfc.nasa.gov/goes/text/geonews.html#FENGYUN>.
 FY-4, 2008. <http://www.spacedaily.com/news/china-01zg.html>.
 GCOM, 2008. http://www.jaxa.jp/projects/sat/gcom/index_e.html.
 GeoEye, 2008. <http://www.geoeye.com/CorpSite/default.aspx>.
 GeoInformatics, 2007. <http://www.geoinformatics.com/asp/default.asp?t=article&newsid=3145>.
 Glackin, D. L., and Peltzer, G. R., 1999. *Civil, Commercial, and International Remote Sensing Systems and Geoprocessing*. El Segundo: The Aerospace Press and AIAA.
 Glory, 2008. <http://glory.giss.nasa.gov/>.
 GOCE, 2008. <http://www.esa.int/esaLP/LPgoce.html>.
 GOES, 2008. <http://www.osd.noaa.gov/GOES/index.htm>.
 GOES-R, 2008. <http://www.goes-r.gov/>.
 GOMS, 2008. <http://rsd.gsfc.nasa.gov/goes/text/geonews.html#GOMS>.
 GPM, 2008. http://www.jaxa.jp/projects/sat/gpm/index_e.html.
 GRACE, 2008. <http://www.csr.utexas.edu/grace/>.
 HSI, 2007. Brazil, Argentina agree to build hyperspectral satellite. *Space News*. November 26, 2007, p. 3.
 HY, 2007. <http://209.85.141.104/search?q=cache:0kocsT4gtlMl:www.ne.jp/asahi/prime-intl/presen/GEOSS/0414/07.pdf+HY-2+Satellite&hl=en&ct=clnk&cd=5&gl=us>.
 Hypseo, 2007. http://www.space.com/spacenews/archive07/carlo_0611.html.
 ICESat, 2008. <http://icesat.gsfc.nasa.gov/>.

- IIPS, 2008. <http://projects.osd.noaa.gov/iips/>.
- IMS-1, 2008. <http://www.thehindu.com/2008/04/28/stories/2008042851471000.htm>.
- India, 2008. <http://www.isro.org/programmes.htm>.
- Jacobsen, 2008. http://209.85.173.104/search?q=cache:FCII7DY6-J0J:www.ipi.uni-hannover.de/uploads/txt_tkpublikationen/HRljac.pdf+SPOT-5+dual-use&hl=en&ct=clnk&cd=3&gl=us.
- Jason, 2008. <http://topex-www.jpl.nasa.gov/mission/jason-1.html>.
- KOMPSAT, 2008. <http://komsat.kari.re.kr/english/index.asp>.
- Kramer, H. J., 2002. *Observation of the Earth and its Environment: Survey of Missions and Sensors*, 4th edn. Berlin: Springer.
- Landsat, 2008. <http://landsat.gsfc.nasa.gov/>.
- LDCM, 2008. <http://ldcm.nasa.gov/>.
- Megha-Tropiques, 2008. <http://meghatropiques.ipsl.polytechnique.fr/>.
- MetOp, 2008. http://www.esa.int/esaEO/SEMLFM2VQUUD_index_0_m.html.
- MisrSat, 2008. http://space.skyrocket.de/index_frame.htm?http://space.skyrocket.de/doc_sdat/egyptsat-1.htm.
- MSG, 2008. http://www.esa.int/esaEO/SEMLFM2VQUUD_index_0_m.html.
- MTSat, 2008. <http://www.jma.go.jp/jma/jma-eng/satellite/index.html>.
- NASA, 2003. A-train fact sheet. www.aqua.nasa.gov/doc/pubs/A-Train_Fact_sheet.pdf.
- NASA, 2007. Remote sensing tutorial. <http://rst.gsfc.nasa.gov/>.
- NPOESS, 2008. <http://www.ipo.noaa.gov/>.
- NPP, 2008. <http://jointmission.gsfc.nasa.gov/>.
- Oceansat, 2008. <http://www.isro.org/irsp4.htm>.
- OCO, 2008. <http://oco.jpl.nasa.gov/>.
- Odin, 2008. http://www.snsb.se/eng_odin_intro.shtml.
- OrbView-2, 2008. <http://oceancolor.gsfc.nasa.gov/SeaWiFS/>.
- PARASOL, 2008. http://smc.cnes.fr/PARASOL/GP_organisation.htm.
- Paz/Ingenio, 2008. Spain spends to secure bigger role in ESA. *Space News*. June 2, 2006, p. 6.
- Pleiades, 2008. <http://smc.cnes.fr/PLEIADES/>.
- POES, 2008. <http://www.oso.noaa.gov/poes/>.
- QuickBird, 2008. <http://www.digitalglobe.com/index.php/85/QuickBird>.
- QuikSCAT, 2008. <http://winds.jpl.nasa.gov/missions/quikscat/index.cfm>.
- RADARSAT, 2008. http://www.ccrs.nrcan.gc.ca/radar/index_e.php.
- Rao, P. K., Holmes, S. J., Anderson, R. K., Winston, J. S., and Lehr, P. E. (eds.), 1990. *Weather Satellites: Systems, Data, and Environmental Applications*. Boston: American Meteorological Society.
- RapidEye, 2008. http://www.rapideye.net/Home/home_frames.htm.
- RASAT, 2006. http://209.85.141.104/search?q=cache:M4kK3PVHIZAJ:www.abgs.gov.tr/tarama/tarama_files/20/SC20DET_SPACE.pdf+RASAT+Satellite&hl=en&ct=clnk&cd=1&gl=us.
- RazakSAT, 2005. <http://209.85.141.104/search?q=cache:DjKjGOCsIHAI:www.aars-acrs.org/acrs/proceeding/ACRS2005/Papers/ANS-5.pdf+RazakSAT+Satellite&hl=en&ct=clnk&cd=1&gl=us>.
- Resourcesat, 2008. <http://www.isro.org/pslvc5/index.html>.
- Resurs, 2008. http://ceos.cnes.fr:8100/cdrom-00/ceos1/satellit/scanex/resurs/resurs_o.htm.
- Resurs History, 2008. <http://www.astronautix.com/craft/resursol.htm>.
- RISAT, 2008. http://directory.eoportal.org/d_ann.php?an_id=12429.
- ROCSAT, 2008. <http://www.nspo.org.tw/2005e/>.
- RSSS, 2008. <http://www.paktribune.com/news/index.shtml?116494>.
- Russia, 2003. <http://209.85.173.104/search?q=cache:ZiFS55Bwmy8J:ceosplenary17.noaa.gov/roshydromet.pdf+Resurs-DK1+satellite&hl=en&ct=clnk&cd=30&gl=us>.
- SAC-C, 2008. http://www.gsfc.nasa.gov/gsfsc/service/gallery/fact_sheets/spacesci/sac-c.htm.
- SAOCOM, 2008. <http://www.invap.net/space/saocom/intro-e.html>.
- SCISAT, 2006. http://www.space.gc.ca/asc/eng/media/news_releases/2006/1006.asp.
- Sentinel, 2008. http://www.esa.int/esaLP/SEMZHMDU8E_LPgmes_0.html.
- SMAP, 2008. <http://smap.jpl.nasa.gov/>.
- SMOS, 2008. http://smc.cnes.fr/SMOS/GP_satellite.htm.
- SSTL, 2008. <http://www.sstl.co.uk/Missions/>.
- TanDEM-X, 2008. http://www.dlr.de/en/desktopdefault.aspx/tabid-1/86_read-4553/.
- Terra, 2008. <http://terra.nasa.gov/>.
- TerraSAR-X, 2008. <http://www.infoterra.de/terrasar-x/terrasar-x-satellite-mission/print.html>.
- THEOS, 2008. <http://www.astrum.eads.net/families/daily-life-benefits/remote-sensing/theos>.
- TOMS/EP, 2008. <http://jwocky.gsfc.nasa.gov/eptoms/ep.html>.
- Triana, 2008. <http://www-pm.larc.nasa.gov/triana.html>.
- TRMM, 2008. http://www.jaxa.jp/projects/sat/trmm/index_e.html.
- Wertz, J. R., 1986. Applied astronomy. *Physics Today*. February 1986, p. 128.
- X-Sat, 2003. http://www.spacedaily.com/reports/PSLV_to_Launch_Singapore_University_Satellite.html.
- YaoGan, 2008. <http://www.nasaspaceflight.com/content/?cid=5111>.
- ZASat, 2008. www.sunspace.co.za/programmes/ZASatPath.htm.
- ZASat-2, 2008. <http://www.ieeexplore.ieee.org/iel5/9116/28908/01303890.pdf?arnumber=1303890>.

Cross-references

Aerosols
 Cloud Properties
 Commercial Remote Sensing
 Cost Benefit Assessment
 Crop Stress
 Cryosphere, Measurements and Applications
 Earth's Radiation Budget: The Top-of-the-Atmosphere Radiation
 International Collaboration
 Lidar Systems
 Lightning
 Mission Costs of Earth-Observing Satellites
 Mission Operations, Science Applications/Requirements
 Observational Platforms, Aircraft, and UAVs
 Observational Systems, Satellite
 Ocean Measurements and Applications, Ocean Color
 Precision Agriculture
 Public-Private Partnerships
 Radar, Scatterometers
 Radar, Synthetic Aperture
 Rainfall
 Remote Sensing, Historical Perspective
 Remote Sensing, Physics, and Techniques
 Sea Surface Salinity
 Sea Surface Wind/Stress Vector
 Severe Storms
 Soil Moisture
 Stratospheric Ozone
 Trace Gases, Stratosphere, and Mesosphere
 Trace Gases, Tropospheric: Detection from Space
 Tropospheric Winds
 Water Resources
 Weather Prediction

OCEAN APPLICATIONS OF INTERFEROMETRIC SAR

Roland Romeiser

Rosenstiel School of Marine and Atmospheric Science,
University of Miami, Miami, FL, USA

Definition

SAR. Synthetic aperture radar – a side-looking imaging radar on a moving platform that exploits the Doppler history of received signals to synthesize the effect of a very long antenna (aperture) with a corresponding high spatial resolution in flight (azimuth) direction.

Interferometry. In this context, the analysis of amplitude and phase differences between corresponding pixels of two complex SAR images of the same scene, acquired from slightly different antenna locations or at slightly different times.

Interferometric SAR (InSAR). A SAR with two (or more) antennas for interferometry.

Along-track interferometry (ATI). Interferometric analysis of SAR images acquired by two antennas with a distance (baseline) in flight direction, which scan the scene at slightly different times. Phase differences are proportional to Doppler shifts and, thus, line-of-sight velocities.

Cross-track interferometry (XTI). Interferometric analysis of SAR images acquired by two antennas with a distance (baseline) perpendicular to the flight direction. Phase differences associated with the cross-track baseline are related to target elevations from a reference plane.

Repeat-pass interferometry. Cross-track interferometry using a single antenna that passes over a test area twice, e.g., during two subsequent satellite repeat cycles. This technique can be used for measuring target elevations and target displacements, e.g., due to an earthquake.

Introduction

When SAR raw data are processed into a SAR image, an amplitude and phase of the backscattered signal can be assigned to each pixel, but the phases in a single image have a uniform distribution and are usually ignored in the data interpretation. However, when two SAR images of the same scene, acquired from slightly different antenna locations (for XTI) and/or at different times (for ATI), are combined, the phase differences between corresponding pixels can contain information on target elevations and/or line-of-sight velocities, respectively. XTI and repeat-pass InSAR techniques are mainly used for land applications and will not be discussed further in the following. In the case of ATI, the phase differences are proportional to the Doppler shift of the signal and thus to line-of-sight target velocities; therefore, an along-track InSAR can be used for high-resolution surface current measurements. This concept was first described by Goldstein and Zebker (1987). For InSARs with combined cross-track and along-track baselines, detected phase variations over water are very similar to those from a pure along-track InSAR, because the surface is basically flat, while phase variations

over land are identical to those from a pure cross-track InSAR as long as there are no moving targets. Siegmund et al. (2004) demonstrated current measurements with such a combined XTI/ATI system. Suchandt et al. (2010) demonstrated speed measurements for moving vehicles on land by spaceborne along-track InSAR.

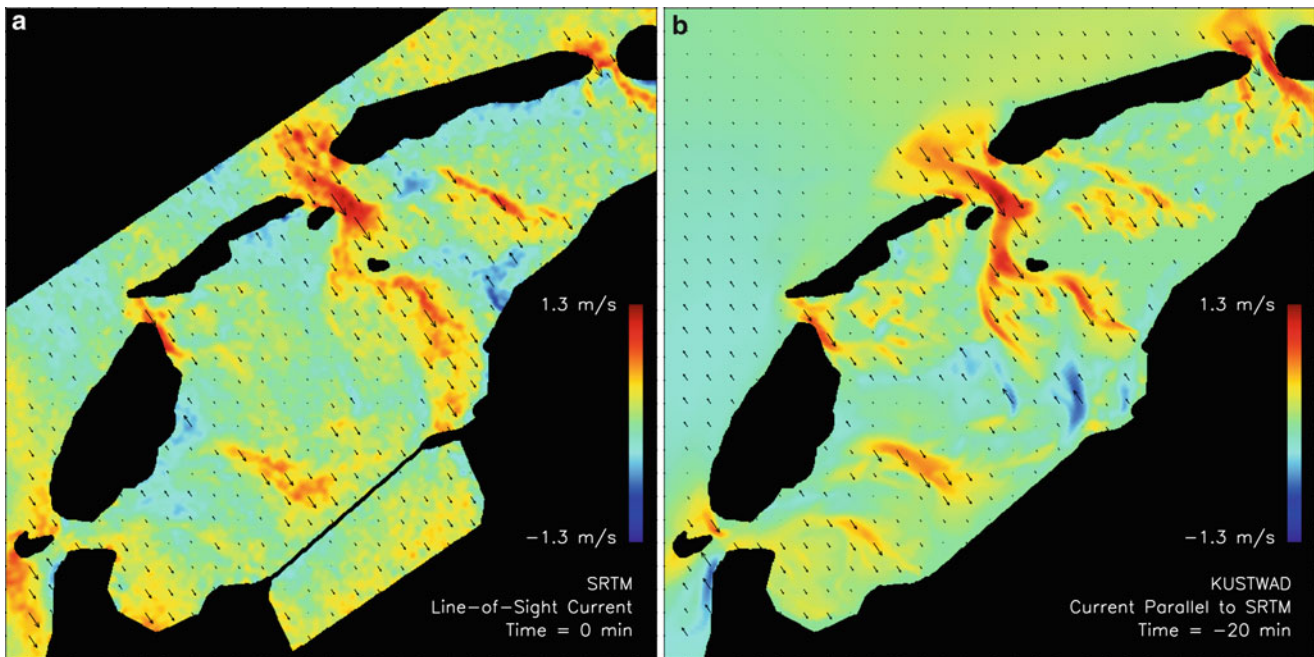
History

Following the ideas formulated by Goldstein and Zebker (1987), the first experiment with an airborne along-track InSAR was carried out over ship-generated internal waves in Loch Linnhe, Scotland, in 1989. Thompson and Jensen (1993) presented results from this experiment and demonstrated that ATI phase signatures include contributions of surface wave motions, which may vary within an image due to hydrodynamic wave-current interaction. Corrections for these contributions could be computed using the theoretical model of Thompson (1989). Graber et al. (1996) showed that properly corrected ATI-derived surface currents from another experiment were in good agreement with HF radar data. Toporkov et al. (2005) demonstrated two-dimensional vector current measurements with an airborne dual-beam ATI system, using two pairs of antennas with different look directions. This technique had been proposed a few years earlier by Frasier and Camps (2001). A first demonstration of current measurements by ATI from space was given by Romeiser et al. (2005), using data from the Shuttle Radar Topography Mission (SRTM) in early 2000 (see Figure 1). The German satellite TerraSAR-X, launched in 2007, is the first civilian satellite with experimental along-track InSAR capabilities in its divided antenna modes; first example results from TerraSAR-X were published by Romeiser et al. (2010).

Data processing

A description of an ATI raw data processing procedure can be found, for example, in the paper by Suchandt et al. (2010). At full resolution, the ATI phases are usually too noisy to show clear signatures of current variations. The phase noise results from instrument noise (dominant at very short time lags), the temporal decorrelation of the backscattered signal (at longer time lags), and spatial fluctuations of the actual Doppler shifts of radar signals received from different patches of the moving water surface. Filters need to be applied to reduce the phase noise and, at the same time, the effective spatial resolution of the data. Examples can be found in the papers by Romeiser (2005) and Romeiser et al. (2005, 2007, 2010).

In the next step, the detected phases have to be corrected for contributions of wave motions. To do this correctly, the wind speed and direction must be known. The wind information can be determined from the radar intensity image itself (Horstmann et al., 2000) or obtained from other sources, such as satellite data, in-situ data, or models. Using the wind vector information and the known radar parameters, a theoretical model, such as the ones



Ocean Applications of Interferometric SAR, Figure 1 (a) Line-of-sight surface current field derived from an interferometric image of the Dutch Wadden Sea from SRTM, acquired on February 15, 2000, 12:34 UTC; area size = $70 \text{ km} \times 70 \text{ km}$; arrows indicate orientation and strength of the current component parallel to the radar look direction; (b) corresponding current field from the numerical circulation model KUSTWAD for the tidal phase 20 min before the SRTM overpass (from Romeiser et al., 2005).

described by Thompson (1989) or Romeiser and Thompson (2000), can be used to estimate the mean contribution of wave motions to the Doppler frequency. A constant correction of this kind was used, for example, by Romeiser et al. (2010). In regions of strong wave-current interaction, spatially varying corrections may need to be computed; a method for doing this has been demonstrated by Romeiser (2005).

If ATI phases are not absolutely calibrated, additional information must be exploited to find a calibration constant to be added or subtracted from the whole array. This was the case with the data from SRTM. Romeiser et al. (2005) determined the calibration constant for an SRTM image of the Dutch Wadden Sea by requiring the current across a dam in the image to become zero. Romeiser et al. (2007) adjusted the calibration constant for an SRTM image of the Elbe river (Germany) in such a way that a uniform flow was obtained for different portions of the river with different flow directions relative to the radar look direction.

Accuracy and spatial resolution

The sensitivity of ATI phases to line-of-sight velocities increases with the ATI time lag, which is determined by the along-track baseline and the platform speed (approx. 7,500 m/s for satellites). At small time lags, the sensitivity is low, and the phase image looks noisy due to instrument noise. At longer time lags, instrument noise becomes less

dominant, but the received signal begins to decorrelate. As shown theoretically by Romeiser and Runge (2007), the best signal-to-noise ratio at X band (the radar frequency band used by SRTM and TerraSAR-X, about 10 GHz, wavelength $\approx 3 \text{ cm}$) should be obtained at time lags of 3–6 ms, corresponding to an effective baseline of 20–45 m. (The actual antenna distance may be twice as long if only one of the two antennas is transmitting and both are receiving.) This optimal time lag or baseline depends on the wind speed (decreasing with increasing wind, since the backscattered signal decorrelates faster) and is roughly proportional to the radar wavelength, i.e., it is an order of magnitude larger for L band (about 1 GHz, wavelength $\approx 30 \text{ cm}$). Even at an optimal time lag, some averaging of detected phases is required to reduce noise, and the effective spatial resolution of noise-filtered data will depend on the time lag and the full-resolution pixel size, as well as the incidence angle and polarization, since the signal-to-noise ratio decreases with increasing incidence angle and it is worse for horizontal (HH) than for vertical (VV) polarization.

For SRTM as well as TerraSAR-X Stripmap/Aperture Switching Mode data, the effective spatial resolution has been found to be on the order of 1 km at an accuracy of 0.1 m/s for measured horizontal currents. This is consistent with the theoretical findings of Romeiser and Runge (2007). With optimized system parameters (longer baseline, reduced instrument noise), the effective resolution could be improved by a factor of about 10. One should

note in this context that there is always a trade-off between the best possible accuracy and spatial resolution, i.e., a better spatial resolution can be obtained at the cost of a reduced accuracy (with less spatial averaging).

Applications

The ATI technique is attractive for all applications that require high-resolution surface current measurements in large areas or at many different locations, where the use of in-situ instruments or ground-based remote sensing techniques (HF radar, nautical radar) is not feasible for technical, economical, or political reasons. Examples are the monitoring of coastal bathymetry (see Calkoen et al., 2001 for a method based on conventional SAR imagery), the siting of turbines for electric power generation from tidal currents, and other coastal and offshore engineering applications. Furthermore, spaceborne along-track InSARs can be useful for a worldwide monitoring of river runoff. Bjerklie et al. (2005) have demonstrated the usefulness of surface current measurements for river runoff estimates with airborne ATI data. Romeiser et al. (2007, 2010) have demonstrated current measurements in rivers by spaceborne ATI. Over the open ocean, satellite-based ATI can be used to monitor oceanic internal waves and current features that cannot be resolved by altimeters, such as small eddies. Airborne ATI systems are useful for applications that require high-resolution current measurements for a limited time, at a higher spatial resolution than an HF radar (or existing satellites) can provide and a better temporal coverage than a satellite can provide (e.g., dedicated flights at certain tidal phases).

Summary

Out of various SAR interferometry techniques, ATI is the most attractive one for applications over coastal area, rivers, or the open ocean. ATI permits direct surface current measurements at spatial resolutions of 1 km and better from a satellite (a factor 10 improvement is possible for future instruments). Since the full-resolution images are noisy and include contributions of wave motions, some filtering and corrections for the wave motions need to be applied; techniques for these processing tasks are readily available. Promising applications are the ones that require high-resolution current measurements in areas where in-situ measurements or the use of other remote sensing techniques is not feasible.

Bibliography

- Bjerklie, D. M., Moller, D., Smith, L. C., and Dingman, S. L., 2005. Estimating discharge in rivers using remotely sensed hydraulic information. *Journal of Hydrology*, **309**, 191–209.
- Calkoen, C. J., Hesselmanns, G. H. F. M., Wensink, G. J., and Vogelzang, J., 2001. The bathymetry assessment system: efficient depth mapping in shallow seas using radar images. *International Journal of Remote Sensing*, **22**, 2973–2998.

- Frasier, S. J., and Camps, A. J., 2001. Dual-beam interferometry for ocean surface current vector mapping. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 401–414.
- Goldstein, R. M., and Zebker, H. A., 1987. Interferometric radar measurement of ocean surface currents. *Nature*, **328**, 707–709.
- Graber, H. C., Thompson, D. R., and Carande, R. E., 1996. Ocean surface features and currents measured with synthetic aperture radar interferometry and HF radar. *Journal of Geophysical Research*, **101**, 25,813–25,832.
- Horstmann, J., Koch, W., Lehner, S., and Tonboe, R., 2000. Wind retrieval over the ocean using synthetic aperture radar with C-band HH polarization. *IEEE Transactions on Geoscience and Remote Sensing*, **38**, 2122–2131.
- Romeiser, R., 2005. Current measurements by airborne along-track InSAR: measuring technique and experimental results. *IEEE Journal of Oceanic Engineering*, **30**, 552–569.
- Romeiser, R., and Runge, H., 2007. Theoretical evaluation of several possible along-track InSAR modes of TerraSAR-X for ocean current measurements. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 21–35.
- Romeiser, R., and Thompson, D. R., 2000. Numerical study on the along-track interferometric radar imaging mechanism of oceanic surface currents. *IEEE Transactions on Geoscience and Remote Sensing*, **38-II**, 446–458.
- Romeiser, R., Breit, H., Eineder, M., Runge, H., Flament, P., de Jong, K., and Vogelzang, J., 2005. Current measurements by SAR along-track interferometry from a space shuttle. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 2315–2324.
- Romeiser, R., Runge, H., Suchandt, S., Sprenger, J., Weilbeer, H., Sohrmann, A., and Stammer, D., 2007. Current measurements in rivers by spaceborne along-track InSAR. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 4019–4030.
- Romeiser, R., Suchandt, S., Runge, H., Steinbrecher, U., and Grünler, S., 2010. First analysis of TerraSAR-X along-track InSAR-derived current fields. *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 820–829.
- Siegmund, R., Bao, M., Lehner, S., and Mayerle, R., 2004. First demonstration of surface currents images by hybrid along- and cross-track interferometric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 511–519.
- Suchandt, S., Runge, H., Breit, H., Steinbrecher, U., Kotenkov, A., and Balss, U., 2010. Automatic extraction of traffic flows using TerraSAR-X along-track interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 807–819.
- Thompson, D. R., 1989. Calculation of microwave Doppler spectra from the ocean surface with a time-dependent composite model. In Komen, G. J., and Oost, W. A. (eds.), *Radar Scattering from Modulated Wind Waves*. Dordrecht: Kluwer Academic Publishers, pp. 27–40.
- Thompson, D. R., and Jensen, J. R., 1993. Synthetic aperture radar interferometry applied to ship-generated waves in the 1989 Loch Linnhe experiment. *Journal of Geophysical Research*, **98**, 10,259–10,269.
- Toporkov, J. V., Perkovic, D., Farquharson, G., Sletten, M. A., and Frasier, S. J., 2005. Sea surface velocity vector retrieval using dual-beam interferometry: first demonstration. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 2494–2502.

Cross-references

Data Processing, SAR Sensors
 Ocean Internal Waves
 Ocean Surface Velocity
 Radar, Altimeters
 SAR-Based Bathymetry

OCEAN DATA TELEMETRY

Michael R. Prior-Jones
British Antarctic Survey, Cambridge, UK

Definitions

Airtime cost. The cost of using a particular communications system. Often charged per minute or per unit of data, but sometimes offered as a flat rate.

Constellation. Collective noun for a group of satellites which function together as a single system.

Data rate. The speed with which data may be transmitted over a communications link. Communications engineers quote this rate in bits/s (or kilobits/s, megabits/s, etc.), whereas many computer systems show data rate in kilobytes/s. One byte is eight bits, so converting from kilobytes to kilobits requires multiplying by eight.

Footprint. The area on the Earth's surface covered by a particular satellite.

Geostationary, geosynchronous, GEO. A satellite with an orbital period equal to the period of the Earth's rotation (the sidereal day, roughly 24 h and 56 min), such that it appears stationary from the Earth's surface. Geostationary orbit is around 35,000 km above the Earth.

ISDN. Integrated Services Digital Network, a telecommunications industry standard for digital communications over telephone lines.

Low Earth Orbit, LEO. A satellite orbiting close to the Earth – usually considered to be between 160 km and 2000 km above the Earth.

Terminal. Equipment used to communicate with a satellite, including the antenna.

Terrestrial radio. Radio systems not involving the use of satellites.

Introduction

Studies of the oceans are increasingly making use of automated instrumentation. These instruments take many forms – such as buoys, moorings, drifting floats, gliders, and autonomous undersea vehicles – but most rely on some form of telemetry to bring their data back for analysis. These data sets are often used in conjunction with remotely sensed measurements, for example, in refining algorithms and assessing the accuracy of the derived variables. A variety of communications systems are available to environmental scientists, many of which are described here. Systems vary in their cost, coverage, data rate, robustness, and power consumption, and so choosing which communications system to use is an important decision in the design of any new autonomous instrument.

Terrestrial radio

Terrestrial radio signals at frequencies of 30 MHz or higher (VHF and up) can be considered as travelling in straight lines. This line-of-sight restriction means that such systems are only suitable for use in coastal observations. A wide range of commercial radio modems are on

the market, typically offering up to 20 km range at up to 10 kbit/s. In many countries terrestrial radio modems may be operated without a radio license, although the law restricts the output power and frequencies that may be used. Licensed operation permits higher output powers and hence longer ranges and/or higher data rates.

Cellular phone systems may also be used. The GSM system used in most countries worldwide has a maximum range of 35 km and can provide data rates of up to 100 kbit/s. Since GSM technology is very small and light, it has been used for tracking the movements of seals (McConnell et al., 2004).

HF radio means frequencies of 3–30 MHz. These frequencies may be reflected from the ionosphere and thus achieve very long ranges (thousands of km) if the propagation conditions are suitable. Due to this reliability issue, the use of HF has largely been eclipsed by satellite communications. However, HF is potentially cheaper than satellite as there is no airtime bill to pay. HF modems using the MIL-STD-188-110B standard are the current state of the art and are available from several defense electronics firms. They offer data rates of around 2400 bits/s.

Geostationary satellite

Geostationary systems are attractive on land because a stationary antenna does not need to track the satellite once the system has been initially set up. At sea, actively tracked antennas are required in order to keep the antenna pointed at the satellite regardless of its movement. Geostationary systems do not provide coverage to the polar regions – their footprints generally extend to 70° north and south.

Inmarsat

Inmarsat's constellation of four satellites covers the whole planet apart from the polar regions. They offer three services aimed at ships (Fleet, FleetBroadband and XpressLink) and two aimed at asset tracking (IsatM2M and IsatDataPro). Inmarsat's other services (such as BGAN, BGAN Link, and BGAN M2M) are designed for use with land-based terminals only and will not work in the marine environment.

Fleet comes in three versions – Fleet 33, 55, and 77. Fleet 33 offers 9.6 kbit/s dial-up data. Fleet 55 offers 64 kbit/s dial-up ISDN. Fleet 77 offers 128 kbit/s dial-up ISDN. All three can also offer an always-on packet-based data service known as MPDS, although this is expensive compared to other products on the market. Fleet terminals consume up to 150 W when transmitting.

FleetBroadband can provide a connection to the Internet at up to 432 kbit/s, although the data rate depends on the terminal used – the smallest, cheapest terminal will only deliver 150 kbit/s. The terminals need around 150 W of power. FleetBroadband airtime costs \$13/MB although bulk users should be able to negotiate a discount.

XpressLink is a hybrid service for ships, consisting of a FleetBroadband terminal and a Ku-band VSAT terminal connected to a common controller. When in range of a Ku-band satellite, the faster Ku-band service is used (up to 768 kbit/s downlink speed), and the system automatically switches over to FleetBroadband (up to 438 kbit/s) when outside the Ku-band area. Inmarsat has a new Ka-band satellite launching in 2014, and XpressLink is upgradeable to support the faster Ka-band service once it becomes available. XpressLink is priced according to bandwidth, not data usage, and prices start from \$2500 per month, including hire of the terminal.

IsatDataPro is a message-based service, supporting messages of a few kilobytes in size (6400 bytes send, 10000 bytes receive). Messages are delivered in around 60 s. There is a monthly subscription that starts at \$19 per terminal, which includes the first 10Kbytes of data. Subsequent data is charged at \$1 per Kbyte. The marine-grade terminal for IsatDataPro is the SkyWave IDP-690, which consumes 0.5 W in receive mode and 9 W in transmit.

IsatM2M is a message-based service, allowing a remote terminal to send messages up to 25 bytes long and receive messages of up to 100 bytes. Messages are delivered within 60s. A typical terminal weighs around 350 g and consumes 6 W on transmit.

Thuraya

Thuraya operates two satellites, providing coverage to Europe, Central and South Asia, and Australia. It covers the Mediterranean, the east Atlantic, the northern part of the Indian Ocean, and the western Pacific.

Thuraya offers dial-up data at 9.6 kbit/s, packet data (GmPRS (1. GmPRS is the name given to Thuraya's packet data service. It's a variant on the GPRS service used on the GSM phone networks)) at 60 kbit/s downlink and 15 kbit/s uplink, plus SMS messaging. All of these services are available from the ThurayaModule terminal, which is about the size of a pack of cards, weighs 60 g and consumes 3 W on transmit.

Thuraya also offers a high-bandwidth product, ThurayaIP, offering up to 444 kbit/s. Using ThurayaIP at sea requires an external stabilized antenna, purchased separately. The entry-level tariff for this service costs \$40/month and then \$6 per megabyte of data transferred.

VSAT

VSAT stands for "very small aperture terminal" which is a rather misleading term. It means the use of commercial telecom satellites (of the kind that are used to carry satellite TV broadcasts) with a relatively small antenna (<3 m diameter). With a VSAT service, you either lease a dedicated channel on the satellite for your exclusive use or join a shared channel operated by a network provider. VSAT is relatively expensive but becomes economic when handling large volumes of data (>100 MB per month).

DCS

The Data Collection Service is provided by a federation of meteorological satellite operators in Europe, the USA, Japan, Russia, and China. Unlike the other systems described, it's free to use. DCS terminals (known as Data Collection Platforms, DCP) normally report hourly or three-hourly and send a single 649 byte message, which takes 1 min to transmit. The system is unidirectional, so the DCP does not know that messages have been successfully received. Messages are delivered within 10 min. DCPs can use as much as 100 W of electrical power during transmission.

An upgraded version of DCS, known as High-Rate DCS (HRDCS), is now in service, and the US GOES satellites are phasing out the old 100 bps service. HRDCS operates at 1200 bps and can transmit the standard 649 byte message in 10 s. Longer messages are possible, up to a maximum of 65536 bytes.

Low earth orbit satellite

Satellites in low Earth orbit move rapidly with respect to the Earth's surface, and so they appear to track across the sky, usually being in view for around 10 min. As the satellites are much closer to the Earth than geostationary orbit, terminals can have smaller antennas and use less transmit power.

ARGOS

ARGOS is a unidirectional, message-based service like DCS, but it uses five polar-orbiting satellites to give global coverage. The message size is limited to 32 bytes, but the terminals can be extremely small and light and are often used for tracking birds, for example. Because there's no acknowledgment from the satellite, the message is repeated several times to try and maximize the chance of good reception.

ARGOS uses the Doppler effect to estimate the position of each transmitter as the message is transmitted. The accuracy of this technique ranges from 350 m to 1 km, depending on how many copies of the message were received by the satellite as it passed overhead. Entry-level ARGOS terminals cost in the region of \$2000, but a considerable premium can be paid for the smallest, lightest models.

A newer version of the system, called ARGOS-3, is being rolled out. At present few satellites support it, but it provides higher data rates (4.8 kbit/s) and two-way communications. The terminals are able to detect the presence of the satellite before it transmits, and to receive acknowledgments for messages received correctly, both of which improve power efficiency.

ARGOS's pricing structure needs some explanation. Rather than charge by the message or by the byte, they charge a monthly fee for each month the terminal transmits plus a fee for each 6 h slot in which messages are sent. Separate tariffs are used for commercial and scientific

Ocean Data Telemetry, Table 1 Comparison of message-based systems

System	Message size	Tariff	Monthly price, 1 message/day	Monthly price, 1 message/h	Terminal power consumption (during transmission)	Two-way comms?	Polar coverage? ^a	Data rate	Time to send one message	Delivery time
Iridium SBD	<340 bytes	\$13/month + \$0.0015/ byte ^b	\$14.24 (30 bytes)	\$31.48 (30 bytes – bulk tariff)	1 W	Yes	Yes		~1 s	<20 s
IsatM2M	25 bytes	\$0.06 for 10 bytes or \$0.120 for 25 bytes	\$5 (25 bytes – minimum spend)	\$89.28 (25 bytes)	9 W	Yes	No		10 s?	30 s
IsatDataPro	<6,400 bytes (tx) <10 kbytes (rx)	\$19/mo + \$1/kbyte ^c	\$19	\$28 (assuming 25 byte message)	9 W	Yes	No		~1 s	<15 s
ARGOS	32 bytes	Land-based fixed stations \$21/month + \$1/6 h slot ^d Other stations \$21/ month + \$1.9/6 h slot	\$54 \$81	\$151	<1 W	No ^e	Yes	480 bps	~1 s	Up to 2 h
DCS	650 bytes	Free	\$0	\$0	50–100 W	No	No	100 bps	75 s	<10 min
Orbcomm	<2,000 bytes?	Prices not available	24 W	Yes	Sporadic	2,400 bps	~1 s	Up to 6 h		
Globalstar simplex	<36 bytes	\$10–\$77/month \$0.2–\$1/msg ^f	\$30 (includes 100 9 byte messages)	\$860	2.5 W	No	No	100 bps	<1 s for 9 bytes, ~3 s for 36 bytes	<30 min

^aPolar coverage means coverage beyond the reach of geostationary satellites (i.e., latitudes higher than 75°)^bThere is a minimum fee per message of \$0.04, covering your first 30 bytes. SBD also has a bulk tariff, where for \$16 a month you get 12,000 inclusive bytes, subject to a minimum bill per message of 10 bytes^cIsatDataPro pricing by OCENS, July 2012. All other prices are from June 2010^dThis is the ARGOS JTA price for scientific applications. Marine animal-tracking devices get a further discount – they are only billed for a maximum of 48 h slots in a given month, regardless of how many they actually use. ARGOS is billed in Euros and the dollar prices here are based on an exchange rate of 1 euro = 1.4 USD^eARGOS currently has one two-way capable ARGOS-3 satellite in orbit, but the ARGOS-3 terminals are not on the market yet^fGlobalstar simplex pricing by Blue Oceans. Cheaper tariffs bill in 9 byte units, and bulk tariff (\$77/month) bills in 36 byte units

applications, and within the scientific tariff, there are different rates for fixed, mobile, and animal-tracking applications.

Iridium

Iridium operates a constellation of 66 satellites providing continuous global coverage and offers voice calls and data services. Although each satellite is only in view for 10 min at most, the network is capable of handing over calls between satellites without losing the connection. They offer three basic data services – Short Burst Data, dial-up data, and OpenPort.

Short Burst Data (SBD) is a bidirectional message-based system. The network supports SBD messages of up to 1960 bytes (2. These figures are for Mobile-Originated SBD messages (i.e., messages from an Iridium terminal). Mobile-Terminated messages (i.e., messages to the Iridium terminal) are slightly shorter Tables 1 and 2), but the lightweight SBD-only modem supports a maximum of 340 bytes per message. Because SBD is

bidirectional, the terminal receives an acknowledgment that its message has been sent successfully. The smallest SBD modem (model 9603) weighs 11 g and consumes ~1 W while transmitting.

Dial-up data offers a 2400bits/s connection in a similar manner to a conventional landline modem. Calls are routed via a modem at the Iridium gateway station and can take up to 40 s to connect. It costs in the region of \$1/min. The 9522B terminal is an off-the-shelf solution for dial-up data – it weighs around 400 g and consumes 4 W while transmitting. There's also a bare-board module (Iridium Core 9523) that weighs 32 g and consumes 3 W while transmitting.

OpenPort (also known as Iridium Pilot) is Iridium's higher-bandwidth product, offering data rates of up to 134 kbit/s. It is intended for use on ships, so the terminal is large and rugged – a cylinder 60 cm in diameter weighing 11 kg and consuming 100 W. OpenPort airtime is priced at around \$7/MB but gets cheaper for higher-volume usage.

Ocean Data Telemetry, Table 2 Comparison of data services

System	Airtime charges ^a Data rate, kbit/s ^c	Monthly airtime cost for... ^b						Polar coverage ^e
		Monthly fee	Charged rate	Equivalent cost per MB ^d	1 MB	10 MB	100 MB	
Iridium dial-up	2.4	\$14	\$1/min	\$58	\$72	\$594	\$5,814	Yes
Iridium OpenPort ^f	9.6	\$0	\$5–\$7/MB	\$5–\$7/MB	\$7	\$68	\$575	Yes
	32	\$0	\$5–\$8/MB	\$5–\$8/MB	\$7	\$72	\$611	
	64	\$0	\$5–\$8/MB	\$5–\$8/MB	\$8	\$76	\$647	
	128	\$0	\$6–\$9/MB	\$6–\$9/MB	\$9	\$84	\$719	
Fleet MPDS	28, 64, 128 ^g	\$0	\$34/MB	\$34	\$34	\$340	\$3,400	No
Fleet 33 dial-up	9.6	\$0	\$3/min	\$43	\$43	\$430	\$4,300	No
Fleet 55/77 ISDN	64	\$0	\$7/min	\$15	\$15	\$150	\$1,500	No
Fleet 77 ISDN2	128	\$0	\$12.50/min	\$14	\$14	\$130	\$1,300	No
FleetBroadband	432 ^h	\$0 ⁱ	\$13/MB	\$13	\$30	\$130	\$1,300	No
Thuraya dial-up	9.6	\$36	\$1/min	\$15	\$51	\$186	\$1,536	No
Thuraya GmPRS	15	\$60 ^j	\$5.50/MB	\$5.50	\$60 ^k	\$88	\$582.5	No
ThurayaIP	444	\$40	\$6/MB	\$6	\$46	\$100	\$585 ^l	No
Globalstar #777	9.6	\$14–\$49	\$1.12/min	\$16	\$30	\$49 ^m	\$49	No

^aAll prices are in US dollars and exclude taxes. Iridium airtime was priced from NAL Research. OpenPort prices were quoted by AST. Fleet prices were from KVH. BGAN, FleetBroadband, Globalstar, and Thuraya (dial-up/GmPRS) prices were from Satphone. ThurayaIP prices were from GTC

^bThis price is the cost per month for the data used in a given month. It includes monthly subscription charges, but does not include initial setup costs such as activation or SIM card fees. Figures have been rounded to the nearest dollar

^cFigures quoted here are uplink speeds – some systems have asymmetric uplink and downlink speeds

^dThis price shows the per-minute rates converted to per megabyte, ignoring monthly fees or any overheads like minutes used while establishing connections

^ePolar coverage means coverage beyond the reach of geostationary satellites (i.e., latitudes higher than 75°)

^fOpenPort allows a choice of data rates, slower is cheaper. It has no compulsory monthly fee, but by paying a monthly rate gets you an inclusive data allowance and a cheaper rate on any excess. Monthly plans start at 10 MB/month

^gMPDS operates at 28 kbit/s on Fleet 33, 64 kbit/s on Fleet 55, and 128 kbit/s on Fleet 77. Airtime prices are the same for all three systems

^hThere are currently two FleetBroadband terminals on the market. The smaller, cheaper unit offers 284 kbit/s data rate. Airtime pricing is the same for both units

ⁱFleetBroadband has no monthly fee, but there is \$30/month minimum spend

^jThe \$60 monthly fee includes the first 5 MB of data allowance

^kIncludes the first 5 MB of data allowance

^lThurayaIP offers 150 MB of data for \$585/month

^mGlobalstar Europe offers an unlimited burst data plan for €34.99 (= \$49). This deal only applies for service within Europe

Globalstar

Globalstar's network architecture means that satellites must be in range of one of its ground stations to operate, so their coverage at sea is limited. Their standard duplex service covers most of the North Atlantic Ocean, the Mediterranean, North Sea, and US Pacific coast. Users purchase a contract for use in a particular region of the world, and roaming to another region is considerably more expensive. Globalstar offers a 9.6 kbit/s dial-up data service and their modem consumes 4 W on transmit.

They also offer a simplex (i.e., unidirectional) message-based data service. This has larger coverage areas and can send up to 36 bytes per message. Simplex modems consume 2.5 W on transmit.

Orbcomm

Orbcomm operates a network of 29 satellites in low Earth orbit, offering a bidirectional message-based communication service. There is no fixed message size, although sending shorter messages is considered more reliable.

Unlike the other services described here, Orbcomm operates in the marine VHF band and hops between frequencies to avoid interference. Coverage is not continuous – holes in the coverage open and close as the satellites move – and there is presently only one polar-orbiting satellite, making polar coverage somewhat erratic.

Messages are downlinked by regional earth stations ("gateways"). If a satellite is in range, the messages are delivered in close to real time. Otherwise, the satellite will store the message and attempt to forward it on when it next comes in range. A terminal is normally associated with just one gateway, and additional fees are payable for "roaming" to other gateways elsewhere in the world.

Orbcomm would not quote airtime prices without a nondisclosure agreement.

Price comparisons

The following tables give indicative prices in US dollars excluding taxes. They are intended to give an estimate as to how much each system costs to operate. As most satellite systems sell airtime through a number of resellers, call more than one supplier in order to get the best price. These prices were obtained in 2010.

Summary

A wide variety of systems are available for data telemetry. Working in coastal waters permits the use of radio modems or cellular phone technology. Satellite technology is widely used for use in remote regions or on the open ocean. Where terminal size and power consumption are all important, a message-based service like Iridium SBD or ARGOS can provide basic communications. Dial-up data (on Iridium, Thuraya, or Globalstar) can provide two-way real-time data comms from compact terminals while still consuming less than 5 W, although airtime is expensive. If plenty of power is available, products like ThurayaIP,

FleetBroadband, and OpenPort provide high data rates at competitive prices for low-volume users. However, for high data volumes (i.e., more than 100 MB per month), VSAT is currently the most cost-effective solution.

Bibliography

McConnell, B., et al., 2004. Phoning home – a new GSM mobile phone telemetry system to collect mark-recapture data. *Marine Mammal Science*, **20**(2), 274–283.

OCEAN INTERNAL WAVES

Werner Alpers

Institute of Oceanography, University of Hamburg,
Hamburg, Germany

Synonyms

Waves in the interior of the ocean

Definition

Ocean internal waves are waves in the interior of the ocean which are generated when the interface between layers of different water densities is disturbed, usually caused by tidal flow over shallow bathymetry.

Introduction

Internal waves are waves of the interior ocean. They can exist when the water body is stratified, i.e., when it consists of water layers of different densities. This difference is usually caused by a difference in water temperature, but it can also be caused by a difference in salinity as in the Strait of Gibraltar. In order to generate internal waves, the interface must be disturbed. There are many ways to disturb the interface, e.g., by (1) water being pushed by the action of the tide over irregularities in shallow bathymetry, such as underwater ridges or shelf breaks (e.g., Lamb, 1994), (2) the transit of a surface ship or a submarine, (3) river outflow, or (4) the passage of an atmospheric front.

Unlike the familiar surface waves, internal waves are not (to first order) associated with an elevation of the sea surface, but with a variable surface current which modulates the sea surface roughness. Remote sensing instruments can capture internal waves via these roughness patterns. The most commonly used instrument for detecting ocean internal waves from space is the synthetic aperture radar (SAR), which is capable of measuring very small changes in the sea surface roughness in the centimeter to decimeter wavelength range. The wave-induced roughness patterns can sometimes even be seen by the naked eye provided that the observer looks at the sea surface with an angle which is close to the specular reflection angle of the sun (into the "sun glint"). Optical sensors flying in space can also be used for detecting internal waves,

but their applicability is restricted to daytime, clear skies, and favorable viewing geometry with respect to the position of the sun (Jackson, 2007).

Basics

When describing internal waves in the ocean, one often approximates the density structure of the water column by two layers of differing densities. Since the change in density is usually due to a change in temperature, one calls this interface the “thermocline.” In the presence of internal waves, the thermocline is distorted. Associated with internal waves are orbital motions of water particles which can move submarines tens of meters up and down. Internal waves as well as surface waves (the familiar waves seen on the ocean surface) are both gravity waves which are governed by the same hydrodynamic equations. The only difference is that in the case of internal waves, the difference in density between the two fluids (warm water above and cold water below) is about three orders of magnitude smaller than in the case of surface waves (air above and water below). This implies that the spatial and temporal scales of ocean internal and surface waves are quite different.

Typical wavelengths of internal waves in the ocean are hundreds of meters to tens of kilometers and typical periods are several minutes to several hours. The orbital motions associated with internal waves generate convergent and divergent flow regimes on the sea surface which

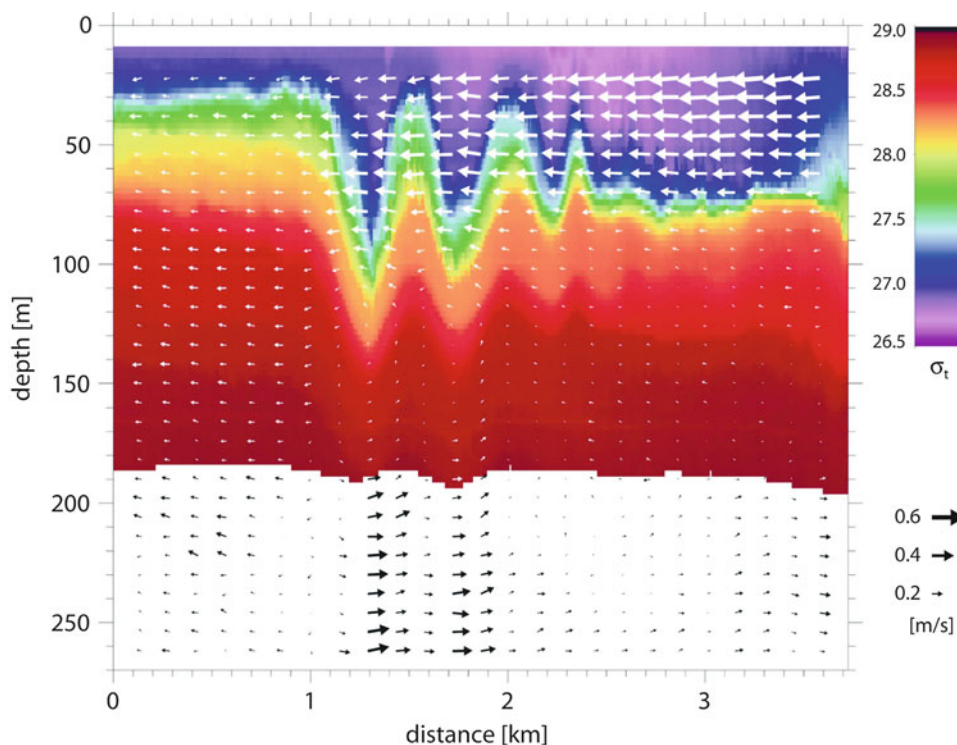
modulate the sea surface roughness by hydrodynamic wave-current interaction (Alpers, 1985). It is by this sea surface roughness pattern that internal waves become detectable by SAR and optical sensors.

Most often ocean internal waves, especially when they are generated by tidal flow over shallow underwater bottom topography, are highly nonlinear and occur in wave packets. The distances between the waves in a wave packet and also their amplitudes decrease from front to rear. Nonlinear internal waves are usually described in terms of soliton theories and are called internal solitons or internal solitary waves (ISWs). The first soliton theory was developed already in 1895 by Korteweg and de Vries.

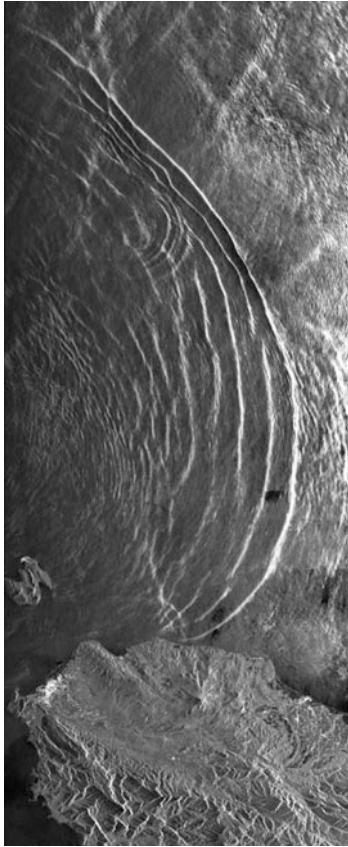
In Figure 1, a typical example of a nonlinear internal wave packet consisting of three ISWs is depicted. The color coding denotes the water density and the arrows the velocity. This vertical profile was measured north of the Strait of Messina in the Mediterranean Sea on October 25, 1995, by a towed conductivity-temperature-depth (CTD) chain and by a vessel-mounted acoustic Doppler current profiler (ADCP). Note that the depression of the thermocline associated with the first solitary wave in the wave packet exceeds 50 m.

Internal waves in the Andaman Sea

The Andaman Sea, located between the Malay Peninsula and the Andaman and Nicobar Islands, is an ocean area



Ocean Internal Waves, Figure 1 Density distribution of the water column and distribution of the velocity north of the Strait of Messina measured by shipborne sensors during the passage of a highly nonlinear internal wave packet on October 25, 1995.



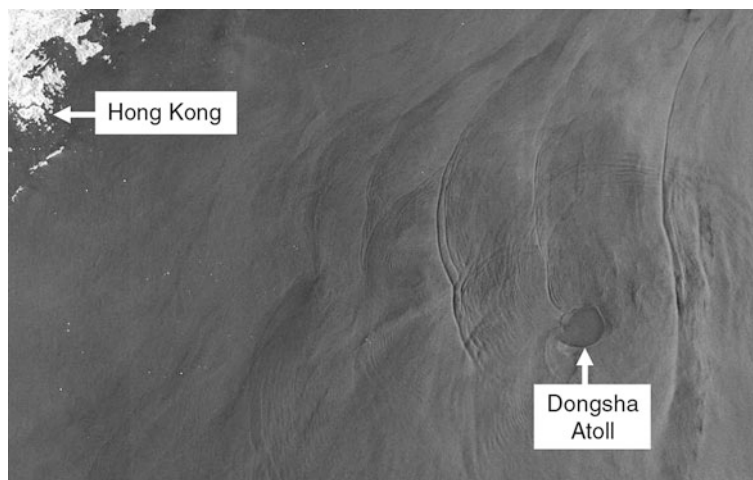
Ocean Internal Waves, Figure 2 ERS-2 SAR image acquired over the Andaman Sea on February 11, 1997, showing sea surface signatures of internal solitary waves. In the lower section of the image, the northwestern part of Sumatra and the Indonesian island Weh are visible. The imaged area is 100×300 km © ESA.

where ISWs of large amplitudes (peak to trough values of more than 60 m) are frequently encountered (Osborne and Burch, 1980). Roughness bands stretching from horizon to horizon generated by ISWs have often been observed in the Andaman Sea from ships. Also, optical images taken from satellites have occasionally revealed sea surface signatures of ISWs. But the ubiquity of large ISWs in the Andaman Sea has become apparent only after a large number of synthetic aperture radar images have become available from the first ESA Remote Sensing satellite ERS-1 launched in 1991 and later from the ERS-2, Radarsat, Envisat, and ALOS satellites launched in 1995, 1995, 2002, and 2006, respectively. Several of the early ERS SAR images of the Andaman Sea can be found in Brandt et al., 1997 and <http://www.ifm.uni-hamburg.de/ers-sar/Sdata/oceanic/intwaves/>

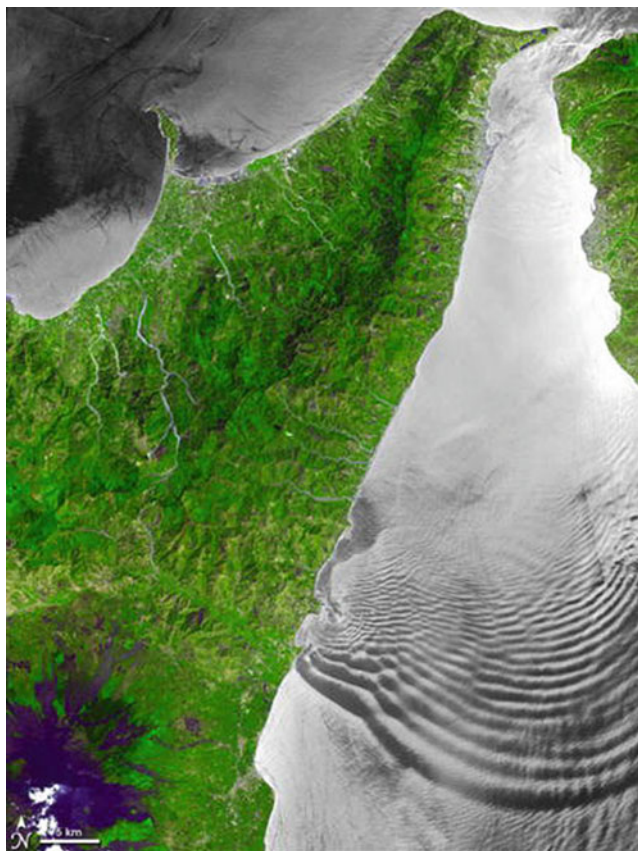
Figure 2 shows an ERS-2 SAR image (swath width: 100 km) which was acquired over the Andaman Sea north of the Indonesian island of Sumatra. The large-scale wave pattern is the sea surface signature of a strong internal solitary wave packet which was generated by the interaction of the tide with shallow bathymetry located between Sumatra and the Great Nicobar Island. Clearly visible are sea surface signatures of six internal solitary waves of decreasing intensity from front to rear. The leading soliton in the wave packet has a crest length of more than 200 km. Also visible in the upper section is a weak semi-circular small-scale wave pattern which is the sea surface signature of a secondary internal wave packet generated by scattering of a large-amplitude internal solitary wave at an underwater bank (Vlasenko and Alpers, 2005).

Internal waves in the South China Sea

The South China Sea is another ocean area where very large-amplitude and long-crested solitary internal waves



Ocean Internal Waves, Figure 3 Envisat ASAR Wide Swath image acquired over the South China Sea showing sea surface signatures of an internal soliton (the long curved line on the right) and several internal wave packets which were generated at successive tidal cycles. The internal waves are refracted by the shallow reef of the Dongsha Atoll. The imaged area is 400×250 km ©ESA.



Ocean Internal Waves, Figure 4 Optical image acquired by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) over the Mediterranean Sea on August 11, 2003. The land surface visible on the *left* is the east coast of Sicily; on the *right*, the west coast of Calabria (Italy); and on the *upper right* corner, the Strait of Messina. Sea surface signatures of a southward propagating internal wave packet generated in the Strait of Messina are visible in the *lower* section of the image (<http://asterweb.jpl.nasa.gov/gallery-detail.asp?name=messina-wave>).

are encountered (e.g., Liu et al., 1998), probably the largest in the world's ocean. They are generated by the interaction of the tidal current with shallow underwater bottom topography located in the Luzon Strait (between Taiwan and Luzon) from where they propagate westward toward the Chinese coast. On their way, they interact with the shallow topography of the Chinese Shelf, in particular with the Dongsha Atoll, and finally dissipate near the Chinese coast. Figure 3 shows an Envisat Advanced SAR Wide Swath image acquired June 18, 2008, at 0213UTC over the South China Sea east of Hong Kong. Visible are sea surface signatures of an internal soliton (the long bended line to the right) and of several internal wave packets generated at successive tidal cycles. Note that the distance between the wave packets decreases from east to west due to the decrease in propagation when the internal waves enter shallow waters.

Internal waves in the Mediterranean Sea

In the Mediterranean Sea, large-amplitude ISWs are generated in the Strait of Gibraltar (separating Spain from Morocco) and in the Strait of Messina (separating the Italian island of Sicily from the Italian Peninsula) by tidal flow over underwater ridges located in these straits. While the existence of internal waves generated in the Strait of Gibraltar had been demonstrated by in situ measurements by 1960, the existence of internal waves was first revealed by a SAR image acquired by the US Seasat satellite on September 15, 1978 (Alpers and Salusti, 1983). On this SAR image, sea surface signatures of a northward propagating wave packet were observed. But in the following years, internal waves propagating northward as well as southward have been detected by in situ measurements and on spaceborne SAR and optical images (e.g., Brandt et al., 1997). An impressive example of an optical image revealing the existence of southward propagating ISWs in the Strait of Messina is depicted in Figure 4. This image was acquired in the visible band by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) onboard the Terra satellite.

Summary

Although ocean internal waves are waves in the interior of the ocean, they are detectable from satellites because they leave a fingerprint on the sea surface by changing the small-scale sea surface roughness. In some cases, these roughness changes can be captured by optical instruments (when the imaged sea area lies in or close to the specular reflection area of the sun), but the best instrument for capturing internal waves from space is the synthetic aperture radar (SAR). SAR images acquired from satellites have greatly contributed to our knowledge of the distribution and dynamics of internal waves in the world's ocean.

Bibliography

- Alpers, W., 1985. Theory of radar imaging of internal waves. *Nature*, **314**, 245.
- Alpers, W., and Salusti, E., 1983. Scylla and Charybdis observed from space. *Journal of Geophysical Research*, **88**, 1800.
- Brandt, P., Rubino, A., Alpers, W., and Backhaus, J. O., 1997. Internal waves in the Strait of Messina studied by a numerical model and synthetic aperture radar images from the ERS 1/2 satellites. *Journal of Physical Oceanography*, **27**, 648–663.
<http://www.ifm.uni-hamburg.de/ers-sar/Sdata/oceanic/intwaves/>
http://www.internalwaveatlas.com/Atlas_index.html
http://www.sarusersmanual.com/ManualPDF/NOAASARManual_CH07_pg189-206.pdf
- Jackson, C., 2007. Internal wave detection using the moderate resolution imaging spectroradiometer (MODIS). *Journal of Geophysical Research*, **112**, C11012, doi:10.1029/2007JC004220.
- Korteweg, D. J., and de Vries, G., 1895. On the change of long waves advancing in a rectangular canal and a new type of long stationary waves. *Philosophical Magazine*, **5**, 422.
- Lamb, K. G., 1994. Numerical experiments of internal wave generation by strong tidal flow across a finite amplitude bank edge. *Journal of Geophysical Research*, **99**, 843.

- Liu, A. K., Chang, Y. S., Hsu, M.-K., and Liang, N. K., 1998. Evolution of nonlinear internal waves in the East and South China Seas. *Journal of Geophysical Research*, **103**, 7995.
- Osborne, A. R., and Burch, T. L., 1980. Internal solitons in the Andaman Sea. *Science*, **208**(4443), 451.
- Vlasenko, V., and Alpers, W., 2005. Generation of secondary internal waves by the interaction of an internal solitary wave with an underwater bank. *Journal of Geophysical Research*, **110**, C02019, doi:10.1029/2004JC002467.

Cross-references

Microwave Surface Scattering and Emission
Observational Systems, Satellite
Radar, Synthetic Aperture
Radars

OCEAN MEASUREMENTS AND APPLICATIONS, OCEAN COLOR

Samantha Lavender
Pixalytics Ltd, Plymouth, Devon, UK

Synonyms

Ocean color

Definition

Ocean color is the spectral variation of the water-leaving radiance that can be related to concentrations of the optically active constituents.

Water-leaving radiance, L_w , is the downwelling solar irradiance that has penetrated the water surface, interacted with the water body (optically active constituents), and then been backscattered towards the sensor used to detect it.

Optically active constituents are those substances dissolved or suspended within the water column that affect the color of the water through absorption and/or scattering.

Ocean color

Introduction and overview of the sensors

The color of water bodies (in this case termed ocean color as this entry will primarily focus on marine waters) was recorded as early as the 1600s; Henry Hudson, sailing from one sea to another while looking for a northeast passage to China and Japan, noted in his ship's log that a sea pestered with ice had a black-blue color (Hudson, 1608). For further information on the historical context (1600–1930), see Wernand (2008).

Spaceborne ocean color imagery was first obtained in the 1960s, through handheld cameras on manned space missions, and clearly demonstrated the potential of satellites for monitoring coastal total suspended matter (TSM) and phytoplankton-induced color variation, e.g., Badgley and Childs (1969). Clarke et al. (1970) undertook the first quantitative work using an airborne

spectroradiometer, recording the influence of different phytoplankton concentrations on the spectral variation of water-leaving radiance.

The first dedicated, but proof of concept, sensor was the Coastal Zone Color Scanner (CZCS) launched onboard the Nimbus-7 spacecraft in 1978. It recorded the water-leaving radiance in four wavebands (430–450, 510–530, 540–560, and 660–680 nm) with a broad near-infrared (NIR) waveband to differentiate sea from land and cloud (700–800 nm) and a thermal infrared band (10.5–12.5 μm) for determining *Sea Surface Temperature (SST)*. However, it was operated on an intermittent schedule (due to power demands), lost the thermal infrared waveband within the first year, and the other wavebands began degrading from 1981 onward although they remained useable to some degree until 1984. These data are still important today and have been reprocessed by both projects and space agencies so that they are as comparable as possible (within the limitations of what was collected) to modern-day datasets. Examples include the North Sea Atlas (Holligan et al., 1989), OCEAN project (Barale et al., 1999), and the NASA Ocean Biology Processing Group reprocessings (see Gregg et al., 2002 and http://oceancolor.gsfc.nasa.gov/CZCS/czcs_processing/).

Since CZCS, a number of ocean color missions have been launched although there was a large gap until the next mission in 1996; the Modular Optoelectronic Scanner (MOS) was designed and built by the German Aerospace Research Establishment (DLR) and launched onboard the Indian IRS-P3 spacecraft (Zimmermann and Neumann, 1997). MOS was launched into a sun-synchronous polar orbit. This is a traditional orbit for ocean color missions as they follow a path that passes close to the North and South Poles so that different parts of the Earth's surface are viewed during successive orbits and the whole of the Earth is covered in 2–3 days if the instrument has a suitable swath width. For MOS, global imagery was not possible as the sensor did not have suitable onboard storage; without this, data has to be directly downlinked while the satellite is in view of a receiving station or has to be relayed to Earth via a data relay satellite, e.g., Artemis, that works alongside Envisat.

The Sea-viewing Wide Field-of-view Sensor (SeaWiFS) has, to date, provided the longest time series of ocean color measurements; it was launched onboard the ORBVUEW-2 satellite in August 1997 (first data in September) and worked until December 2010 although there were several data gaps due to, e.g., problems with a telemetry collection anomaly. Therefore the SeaWiFS dataset is over 10 years in length, which considerably exceeds the original planned lifetime of 5 years. SeaWiFS was a commercial satellite, owned by GeoEye, but with the development funded by the National Aeronautics and Space Administration (NASA) which purchased (alongside agencies such as National Oceanic and Atmospheric Administration, NOAA, in later years) the right to use SeaWiFS data for academic research from

the commercial owner. From 2005, noncommercial data were limited to just the global area coverage (GAC; 4 km sampling) rather than the higher-resolution (approximately 1 km) local area coverage (LAC) measurements that were previously acquired by a number of receiving stations around the world as well as having a small amount recorded onboard. However, in 2008, NASA reactivated the opportunity to receive LAC data in a non-near-real-time (NRT) mode for United States (US) coastal waters; there was always an opportunity to purchase a commercial license. SeaWiFS had eight narrow (20–40 nm) bands in the visible and NIR (402–422, 433–453, 480–500, 500–520, 545–565, 660–680, 745–785, and 845–885 nm), but no thermal waveband for measuring Sea Surface Temperature (SST).

In the twenty-first century, SeaWiFS was joined by the global coverage Moderate Resolution Imaging Spectrometer (MODIS) and Medium Resolution Imaging Spectrometer (MERIS). The MODIS sensors were launched by NASA onboard both the EOS Terra satellite in December 1999 and the EOS Aqua satellite in May 2002. MODIS has additional ocean color wavebands compared to SeaWiFS totaling nine (405–420, 438–448, 483–493, 526–536, 546–556, 662–672, 673–683, 743–753, and 862–877 nm) plus visible and NIR wavebands for atmospheric and terrestrial applications; of particular interest is a narrow waveband for solar-stimulated chlorophyll fluorescence and additional thermal bands for SST.

MERIS was launched onboard the Envisat satellite, by the European Space Agency (ESA), in March 2002, and carried a total of nine instruments that include the Advanced Along-Track Scanning Radiometer (AATSR) instrument for SST applications; the Envisat satellite, and hence MERIS dataset, ceased operating at the beginning of April 2012. MERIS had narrow and selectable wavebands; the standard setup has been 407.5–417.5, 437.5–447.7, 485–495, 505–515, 555–565, 615–625, 660–670, 677.5–685, 700–710, 750.0–757.5, 758.75–761.25, 770–780, 855–875, 885–895, and 895–905 nm that included a solar-stimulated chlorophyll fluorescence waveband. It had a full spatial resolution of 300 m for coastal areas and a reduced resolution of approximately 1 km with global coverage every 2–3 days (Rast et al., 1999); the full-resolution data were available over the whole of Europe (because of Artemis) and where there were receiving stations worldwide. The MODIS ocean color wavebands have a resolution of approximately 1 km, but the terrestrial wavebands have a higher spatial resolution and so the wavebands at 500 and 250 m resolution have been used for coastal applications (e.g., Miller and McKee, 2004 and Shutler et al., 2007). Both MODIS and MERIS data suffer from sunglint contamination as, unlike SeaWiFS, they do not have the ability to tilt away from the sun glitter. The coverage of MERIS was further reduced by the relatively narrow swath (1,150 km as compared to 2,330 km for MODIS).

The Visible Infrared Imaging Radiometer Suite (VIIRS), which is a multispectral scanning radiometer with 22 wavebands between 0.4 and 12 μm , was launched in October 2011 on the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP) satellite, satellite renamed to the Suomi National Polar-orbiting Partnership or Suomi NPP. Although NASA provided assistance with the prelaunch preparations and is distributing processed data, the mission is operated by NOAA as it is seen as an operational mission.

Europe plans to launch the Ocean & Land Colour Instrument (OLCI) on Sentinel-3 as part the Global Monitoring for Environment and Security (GMES) initiative. This sensor is of a similar design and specification to MERIS with an additional channel at 1.02 μm to enhance the existing MERIS atmospheric and aerosol correction capabilities. The first Sentinel-3 is planned for launch in 2014 with follow-up launches to meet observational requirements, such as improved global coverage, and to enable robust and continuous data provision.

Theory behind ocean color and its processing to biogeochemical products

The successful exploitation of ocean color requires a conversion of the top of atmosphere radiances (as detected by the sensor) to water-leaving radiances and then derivation of the optically active biogeochemical constituents such as (see Mueller et al., 2003 for details and references of the in situ determination techniques):

- TSM: It is also referred to as Suspended Particulate Matter (SPM). It represents the total mass of suspended particles that can be broken down into the inorganic and organic fractions. The organic fraction includes the weight of phytoplankton, and the in situ laboratory method for determining the TSM concentration involves filtering water samples through filter papers and then drying and weighing them to determine the added weight.
- Colored dissolved organic matter (CDOM): It is also called gelbstoff or yellow substances or gilvin and consists of humic and fulvic compounds of either terrestrial origin (are transported into marine waters by river/estuary systems) or those resulting from the breakdown of phytoplankton. The concentration is typically measured in a spectrophotometer as an absorption at a specific wavelength and slope of the exponential curve between two wavelengths. This can be combined with the organic matter absorption associated with TSM and called the colored dissolved and detrital organic materials (CDM).
- Chlorophyll-a (Chl-a): It is used to represent the phytoplankton biomass and an input to products such as primary production (PP). Chl-a is often the primary pigment in phytoplankton, but there are several other pigments and together these form the pigment

absorption spectra and are increasingly used in understanding phytoplankton functional types (Nair et al., 2008), e.g., phycoerythrin and phycocyanin (phycobilins) for cyanobacteria. Determining Chl-a in the laboratory or on ships can be via fluorescence or absorption, but the increasingly preferred method that also provides a full range of pigments is high-performance liquid chromatography.

Before the methods for determining the biogeochemical constituents are applied, an atmospheric correction (AC) must be performed as only 5–10 % of the total signal originates from the water. For the basic AC, the total reflectance at the top of the atmosphere can be written as

$$\rho_t(\lambda) = \rho_r(\lambda) + \rho_a(\lambda) + \rho_g(\lambda) + t \cdot \rho_w(\lambda) \quad (1)$$

where $\rho_r(\lambda)$ is the Rayleigh scattering reflectance, $\rho_a(\lambda)$ is the aerosol scattering reflectance, and $\rho_g(\lambda)$ is glitter from the water surface; see Figure 1 for the relative contributions of the various components in Equation 1. The diffuse atmospheric transmission, t , can be approximated as (Gordon et al., 1983)

$$t = \exp[-(0.5\tau_r + \tau_{oz})/\cos\theta_v] \quad (2)$$

where τ_r is the Rayleigh optical thickness, τ_{oz} is the ozone optical thickness, and $\cos\theta_v$ is an approximation of the path length (cosine of the satellite zenith angle).

In Case I waters (colored by biogenic materials alone, i.e., phytoplankton, its pigments, dissolved organic exudates, and detritus), the term $t \cdot \rho_w(\lambda_{NIR})$ approximates zero and so with the term $\rho_r(\lambda_{NIR})$ being calculated directly (a function of the path length and atmospheric pressure) then can be determined given two NIR wavebands: extrapolation of $\rho_a(\lambda)$ using either the Angstrom exponent, n , or a variable c given in Gordon and Wang (1994). From this point onward, only the Angstrom exponent will be considered.

The NIR aerosol reflectance ratio is

$$\varepsilon_a(\lambda_{NIR(1)}, \lambda_{NIR(2)}) = \rho_a \frac{(\lambda_{NIR(1)})}{\rho_a(\lambda_{NIR(2)})} \quad (3)$$

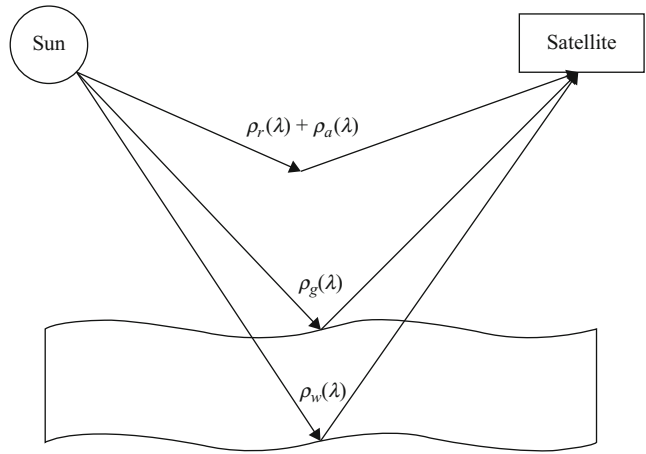
which is used to calculate Angstrom exponent:

$$n = \ln \frac{[\varepsilon_a(\lambda_{NIR(1)}, \lambda_{NIR(2)})]}{\ln \left[\frac{\lambda_{NIR(1)}}{\lambda_{NIR(2)}} \right]} \quad (4)$$

Then, $\rho_w(\lambda)$ is calculated as

$$\rho_w(\lambda) = \frac{\left[\rho_t(\lambda) - \rho_r(\lambda) - \rho_{as}(\lambda_{NIR(2)}) \cdot \left(\frac{\lambda}{\lambda_{NIR(2)}} \right)^n \right]}{t} \quad (5)$$

However, in Case II waters (major influence is TSM and/or CDOM), $\rho_w(\lambda_{NIR})$ is no longer zero and the observed $\varepsilon_a(\lambda_{NIR(1)}, \lambda_{NIR(2)})$ becomes



Ocean Measurements and Applications, Ocean Color,

Figure 1 Simplified diagram of the relative contributions to the signal received by a satellite-based sensor looking at a water surface as given in Equation 1.

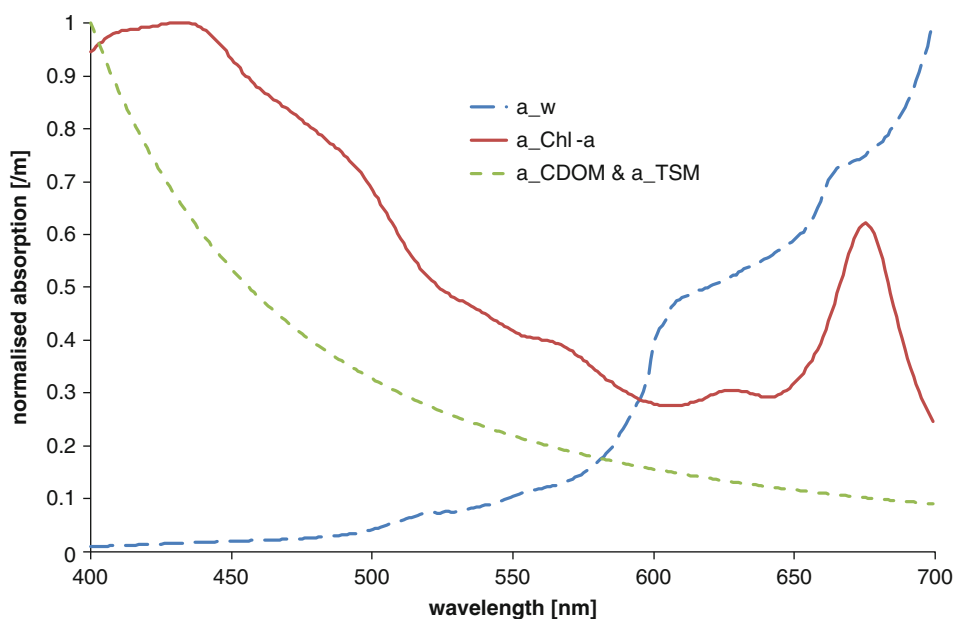
$$\varepsilon_a(\lambda_{NIR(1)}, \lambda_{NIR(2)}) = \frac{[\rho_{as}(\lambda_{NIR(1)}) + t \cdot \rho_w(\lambda_{NIR(1)})]}{[\rho_{as}(\lambda_{NIR(2)}) + t \cdot \rho_w(\lambda_{NIR(2)})]} \quad (6)$$

In order to solve this problem, it is necessary to develop a coupled hydrological and atmospheric optical model in the NIR (700–900 nm) that provides estimates of $\rho_a(\lambda)$, often termed the bright pixel atmospheric correction as the waters are bright in the NIR. SeaWiFS initially used an iterative method (Siegel et al., 2000) to correct for non-negligible water reflectance arising from moderate to high phytoplankton abundances (chlorophyll concentrations greater than $\sim 2 \text{ mg m}^{-3}$) and so independent research was applied to correct for the effects of TSM (e.g., Ruddick et al., 2000 and Lavender et al., 2005), but the processing software was then updated so that it included non-phytoplankton scattering (Arnone et al., 1998), applied to both MODIS and SeaWiFS (Stumpf et al., 2002). For MERIS, the algorithm of Antoine and Morel (1999) was implemented as the AC with the additional bright pixel correction of Moore et al. (1999) implemented from launch. In addition, research has also been done using the longer shortwave infrared (SWIR) wavelengths, where these are available, as the water signal can still be assumed to be negligible (Wang and Shi, 2005).

Once the AC has been performed, the above-water reflectance, $\rho_w(\lambda)$, can then be used to derive the biogeochemical constituents using a hydrological optics model (Gordon et al., 1975; Morel and Prieur, 1977):

$$\rho_w(\lambda) = f \left[\frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \right] \quad (7)$$

where f is an interface term defined according to Morel and Gentili (1993) that approximately equals 0.33, b_b is



Ocean Measurements and Applications, Ocean Color, Figure 2 Normalized specific absorption for water and the different biogeochemical constituents (Chl-a and CDOM/TSM).

the total backscattering, and a is the total absorption. In turn, the total backscattering and absorption are the sum of the biogeochemical constituent components:

$$a = a_w + SPM \cdot a_{SPM} + a_{CDOM} + a_{ChI-a} \quad (8)$$

$$b_b = 0.5b_w + SPM \cdot b_{bSPM} + b_{bCDOM} + b_{bChI-a} \quad (9)$$

The factor of 0.5 in the water scattering term of Equation 9 converts the scattering values to backscattering values and can be applied because of the symmetry of the volume scattering function (Morel, 1974). The chlorophyll absorption and backscattering terms do not have concentration terms in Equations 8 and 9 because they have nonlinear relationships with concentration. Equation 9 also describes a monodisperse (no multiple scattering) system with particles of a given refractive index scattering light, i.e., the scattered light can be assumed to have the same wavelength as the incident light and the particles are independent (i.e., the intensities of the light scattered by each particle can be summed so that the total scattering is proportional to the number of particles). Figure 2 shows the spectral shapes of the normalized specific absorption for water and the different biogeochemical constituents (Chl-a and TSM/CDOM).

For SeaWiFS, instead of reflectance, the normalized water-leaving radiance has been used (i.e., the water-leaving radiance that would be measured by a sensor placed above the water surface if the sun were at the zenith and the atmosphere were absent). The full equation is given in Gordon et al. (1988) and it can easily be converted to a reflectance by dividing it by the solar irradiance.

One method of determining the concentrations of the optically active biogeochemical constituents is to rearrange Equation 7 so that the input is the spectral reflectance and the output is the total Inherent Optical Properties (IOPs), which can subsequently be split into the individual concentrations using Equations 8 and 9. Solving the equations can be performed in a number of ways with various assumptions being made, and so there are a number of IOP models with examples including the Carder model (Carder et al., 2003), the epsilon model (Pinkerton et al., 2006; Smyth et al., 2006), the GSM01 model (Garver and Siegel, 1997 and Maritorena et al., 2002), and the quasi-analytic algorithm (QAA) model (Lee et al., 2002). However, this diversity has caused problems in picking the “best” model. Therefore, algorithms have tended to use the empirical rather semi-analytical approach. There is also the analytical approach, but this is difficult to take forward because of the further work needed in measuring the specific and bulk IOPs, specifically the backscattering coefficient, as discussed in Morel and Maritorena (2001).

To recap, in Case 1 waters the reflectance primarily depends on the Chl-a concentration. This is not linear because of an interdependence that will not be described further here, but the concentration range varies from very low values ($\sim 0.02 \text{ mg/m}^3$) in the oligotrophic areas (such as the open ocean gyres where the chlorophyll maximum is at depths of 150–200 m) up to high values (typically $5\text{--}20 \text{ mg/m}^3$) in upwelling areas and bloom events in coastal areas and the mid to high latitudes. For SeaWiFS and MODIS, the empirical approach is employed with the standard Chl-a product for SeaWiFS, for example, the OC4v4 chlorophyll algorithm

(O'Reilly et al., 1998, 2000), as chosen during the SeaWiFS Bio-optical Algorithm Mini-workshop (SeaBAM) and shown in Equations 10 and 11. For MERIS, the semi-analytical approach is used with a maximum band-ratio algorithm in Case I waters (MERIS, 2007) and an Inverse Radiative Transfer Model-Neural Network (IRTM-NN) in Case II waters (Schiller and Doerffer, 1999 and Doerffer and Schiller, 2000). A detailed overview of these different approaches is given in IOCCG (2006).

$$\text{Chl-a} = 100^{0.366 - 3.067x + 1.930x^2 + 0.649x^3 + 1.532x^4} \quad (10)$$

$$x = \log_{10} \left(\frac{R_{rs443} > R_{rs490} > R_{rs510}}{R_{rs555}} \right) \quad (11)$$

Increasingly important in remote sensing is merging of the data from different sensors to produce multisensor products. Individual sensors typically have a predicted (planned for) lifetime of 5 years, but in reality they can have lifetimes that are much shorter (e.g., the Ocean Color Temperature Sensor (OCTS) onboard the ADEOS platform lasted for only 11 months due to a technical fault) and longer (SeaWiFS and MERIS collected data for over a decade as previously mentioned). Several merged ocean color datasets have become available in recent years:

- NASA OBPB merged Chl-a dataset based on MODIS-Aqua and SeaWiFS
- NASA REASoN merged Chl-a, CDM, and particulate backscattering (b_{bp}) datasets based on MODIS-Aqua and SeaWiFS (e.g., Maritorena and Siegel, 2005)
- ESA DUE GlobColour multiparameter (19 in total) dataset based on MERIS, MODIS-Aqua, and SeaWiFS (Pinnock et al., 2008)

The merging not only provides a potentially longer time series, but also increases the temporal coverage and error characterization of the final dataset. For SeaWiFS, MODIS-Aqua, and MERIS, the average daily cover is between ~8 % (MERIS) and ~16 % (SeaWiFS); see IOCCG (2008). When pairs of sensors are combined, the daily coverage reaches the ~20–25 % range, while the combination of the three sensors allows about 30 % of the ocean to be covered daily (IOCCG, 1999). Similar improvements in coverage occur for 4 day composites with a maximum of about 60 % ocean coverage with three sensors (Gregg et al., 1998). The main causes for loss of coverage are clouds and sunglint.

Applications

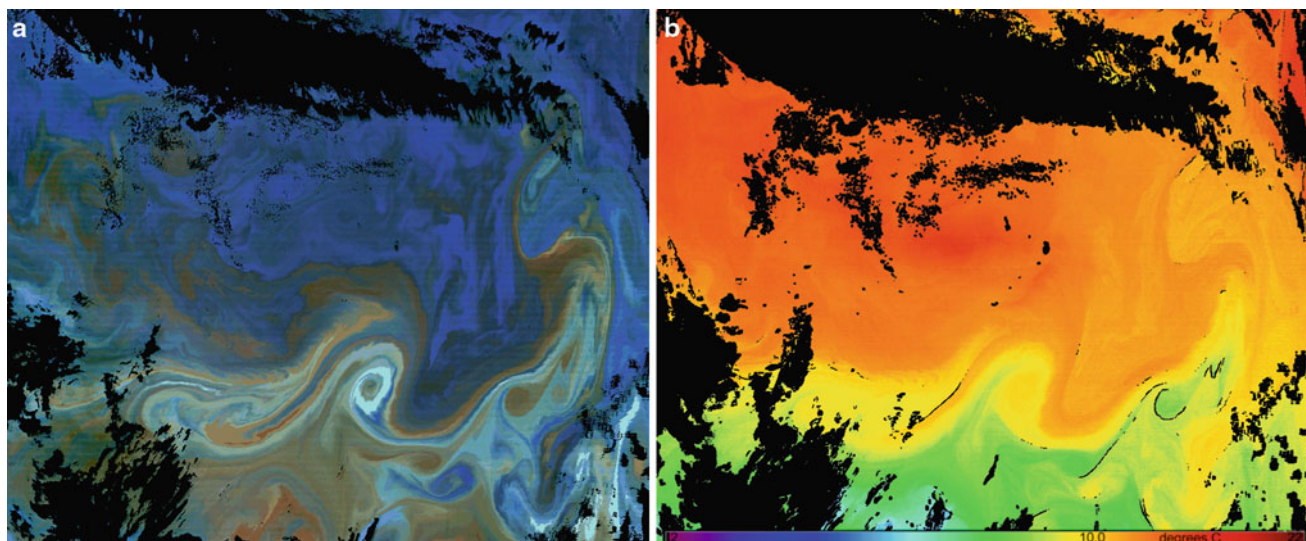
This section highlights some example applications from open ocean to coastal waters and also the linkage with satellite/model sources of nonoptical data. A more complete review of the applications of satellite ocean color data is given within IOCCG report 7 titled “Why Ocean Colour?” (IOCCG, 2008).

SST has previously been mentioned and it is often used together with ocean color data for biological applications;

for an overview of SST, see “*Sea Surface Temperature*.” Increased Chl-a levels and variations in SST are frequently found along frontal zones, within upwelling regions and physical features, e.g., eddies, where zooplankton and fish populations are known to accumulate for early life development spawning and feeding. Figure 3 shows an example MODIS image from December 26, 2006, for an area north of South Georgia where the Antarctic Circumpolar Current, the Subantarctic Front, and the Polar Front produce eddies that can be seen in both the pseudo true-color and SST imagery. A paper analyzing SeaWiFS imagery in this area is Korb et al. (2004) with fisheries discussed further in “*Fisheries*.”

In addition, PP can be derived from the combination of ocean color and SST and has previously been linked to fish yields and coastal upwelling regions (e.g., Carr, 2001; Simpson, 1994, and Platt and Sathyendranath, 2008). The simplest PP models are depth integrated and based upon Chl-a or the product of depth-integrated Chl-a and daily integrated surface Photosynthetically Available Radiation, PAR (Behrenfeld and Falkowski, 1997). The more complex versions bring in SST and an understanding of plankton physiology that characterizes light absorption (Platt, 1986). Campbell et al. (2002) compared a selection of satellite PP algorithms developed by several teams and concluded that the level of agreement between the algorithms had no apparent relationship to the mathematical structure or complexity of the algorithms. A further comparison in 2006 (Carr et al., 2006) that also included biogeochemical ocean general circulation models (BOGCMs) concluded that the groups did not follow model complexity with regard to wavelength or depth dependence, though they were related to the manner in which temperature was used to parameterize photosynthesis. Carr et al. (2006) summarized that further progress in modeling PP required an improved understanding of the effect of temperature on photosynthesis and better parameterization of the maximum photosynthetic rate. Friedrichs et al. (2009) compared satellite-based depth-integrated PP algorithms and BOGCMs to a tropical Pacific PP in situ database, concluding that interdecadal and global changes will be a significant challenge for both EO PP algorithms and BOGCMs.

As previously mentioned, MODIS and MERIS also have a relatively narrow waveband at around 682.5 nm to detect the solar-stimulated chlorophyll fluorescence peak (Rast et al., 1999 and Gower and Borstad, 2004), which is closely linked to cell physiology and can vary with nutrient status, species composition, and growth rate. Behrenfeld et al. (2009) suggest satellite fluorescence is a valuable tool for evaluating nutrient stress predictions in ocean ecosystem models and provide the first synoptic observational evidence that iron plays an important role in seasonal phytoplankton dynamics of the Indian Ocean. In addition, research has extended this fluorescence waveband to applications such as floating sargassum seaweed (Gower et al., 2006). PAR can be derived from ocean color data (Frouin et al., 1989 and Frouin et al.,



Ocean Measurements and Applications, Ocean Color, Figure 3 MODIS-Aqua Level-2 image from December 26, 2006, showing the area north of South Georgia as (a) pseudo true-color composite with 667 nm as red, 531 nm as green, and 443 nm as blue and (b) SST with a rainbow color palette applied. Black indicates the presence of clouds.

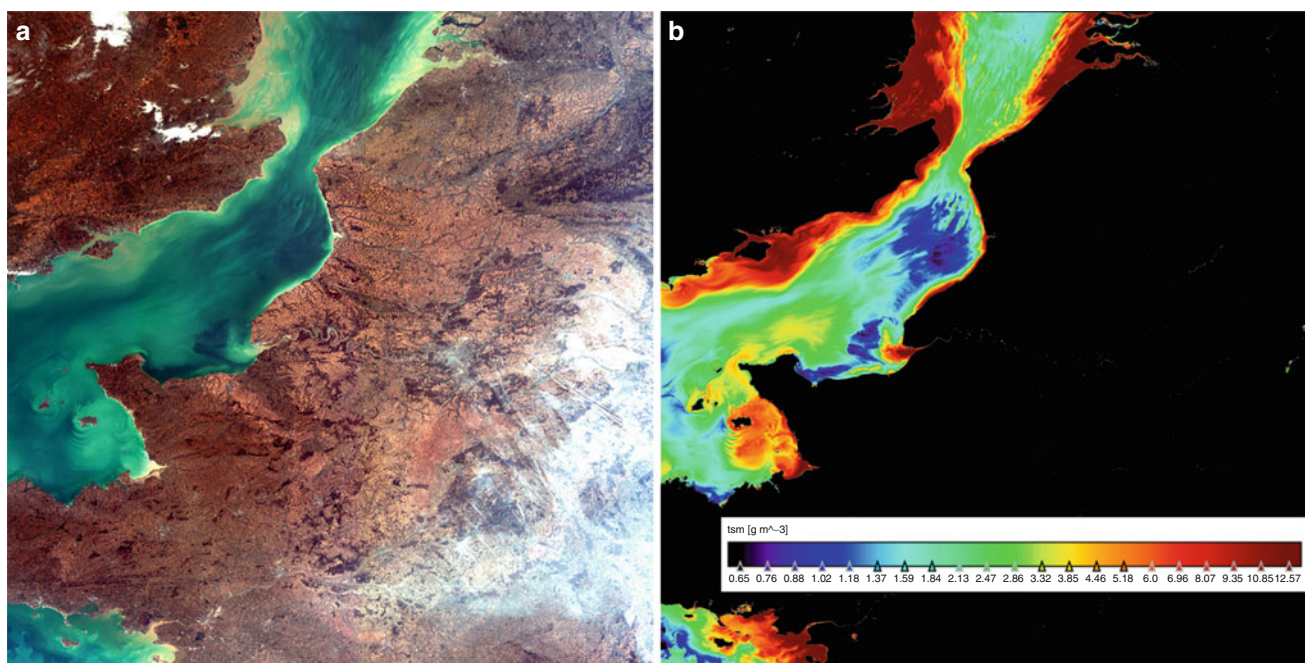
2005) and is a useful parameter because it can act as a limiting factor for bloom initiation/growth, e.g., Raitso et al. (2006) described coccolithophore bloom size variation in response to the regional environment of the subarctic North Atlantic and showed that the combined effect of high solar radiation (PAR), shallow mixed layer depth (derived from a 3D physical model), and increased temperatures (SST) was highly correlated with abundance.

Harmful algae blooms have commonly been called red or brown tides as it has been assumed the red or brown coloration is given to the water by high concentrations of phytoplankton such as dinoflagellates. However, Dierrsen et al. (2006) showed that the color shift, from green to red, comes from an overlap in spectral response of the human eye's red and green cones (centered at 564 and 534 nm, respectively) rather than absorption by the phytoplankton. As shellfish filter large quantities of water, they concentrate the phytoplankton in their tissues, which can affect the aquaculture industries directly by mass fish mortalities and indirectly by shellfish accumulating toxins that can, under certain circumstances, render the shellfish unsafe for human consumption. The monitoring of a bloom (once it has been detected) is important, but it would also be desirable to forecast its occurrence from spectral quality, thermal signature, and hydrography (Tester and Stumpf, 1998). The spectral information can be used to identify different groupings of phytoplankton if they are in abundance and have characteristic absorption/scattering properties, and hence provide an assessment of the potential risks; functional type information is an active area of research (e.g., Sathyendranath et al., 2004; Alvain et al., 2005, 2008; Nair et al., 2008; Raitso et al., 2008). An understanding of biological variability may also be enhanced with additional satellite data

sources such as wind stress (e.g., Pradhan et al., 2006, derived from the second European Remote-Sensing Satellite, ERS-2, and NASA QuikSCAT mean wind fields) and radar altimetry Sea Surface Height (SSH) that links to biophysical coupling (e.g., Wilson and Adamec, 2002).

Water quality is of paramount importance to managers in coastal waters, especially in Europe with the implementation of the EU Water Framework Directive (WFD) that came into force in 2000. Traditional water quality monitoring depends on in situ measurements and the subsequent laboratory analysis of the samples with satellite remote sensing, allowing for the integration and extrapolation of these measurements. The optical remote sensing techniques and products detailed here apply, but these waters cause particular remote sensing challenges as they are a complex mixture of optically active constituents, exhibit significant heterogeneous patterns, and may only be a few pixels in an ocean color satellite image with a significant number of "edge" pixels contaminated by the brighter land signal. Therefore, this community has also relied on land-orientated satellites and airborne remote sensing, e.g., Doxaran et al. (2002). Figure 4 shows a MERIS full-resolution (300 m) Level-2 image from February 11, 2008, showing the southern UK, English Channel, and northern France. Both the pseudo true-color composite and TSM image (processed using the MERIS Case II Regional Neural Network) show the TSM variations with a particular feature around the Channel Islands.

Since the 1980s, inland water monitoring has tended to use the statistical correlation between broad-waveband reflectances and optically active constituents (Chl-a and other pigment concentrations such as Phycocyanin within



Ocean Measurements and Applications, Ocean Color, Figure 4 MERIS full-resolution Level-2 image from February 11, 2008, showing the southern UK, English Channel, and northern France as (a) pseudo true-color composite with 665 nm as red, 510 nm as green, and 443 nm as blue and (b) TSM with a rainbow color palette applied.

cyanobacterial blooms, Secchi disk depth, and TSM) of the water column (e.g., Kloiber et al., 2002; Vincent et al., 2004). Many studies have also used satellite imagery to map the surface temperature of freshwater inland lakes (e.g., Li et al., 2001; Oesch et al., 2005), and important future variables will also include water levels derived from radar altimetry (e.g., Frappart et al., 2006; Berry et al., 2011).

In nearshore coastal waters, the water depth may be sufficiently shallow that the light penetrates to and returns from the bottom allowing users to map bathymetry, habitats, and vegetation. Mumby et al. (2001) demonstrated the use of fine (1 m) spatial resolution airborne multispectral imagery. Lubin et al. (2001) explored the suitability of existing sensors such as Landsat, while Mumby and Edwards (2003) examined the cost and accuracy of IKONOS as compared directly to a suite of satellite and airborne instruments including CASI and Landsat and SPOT sensors.

Future global or geostationary hyperspectral satellite sensors (with increased spectral resolution) would aid our ability to remotely sense rapidly changing nearshore coastal environments and undersampled inland waters. ESA's Compact High Resolution Imaging Spectrometer (CHRIS), launched in October 2001 onboard the Project for On-Board Autonomy small satellite (Cutter et al., 2003; Barnsley et al., 2004), was not designed as an ocean color mission. However, its multiangle (fly-by zenith angles of 0 , ± 36 , and $\pm 55^\circ$) and hyperspectral (although the users have defined modes with a selection of 18–62 wavebands and 17–34 m spatial resolution) capabilities have spawned

inland and coastal applications (Lavender et al., 2005; Sterckx et al., 2005; Van Mol and Ruddick, 2005). In addition, Neukermans et al. (2009) undertook a feasibility study with SEVIRI where the determination of TSM was assessed. In 2010 the South Korean launched the first ocean color sensor into a geostationary orbit.

Summary

This entry started by reviewing the history of ocean color and then moved on to the theory and processing with a final section on the applications. The opportunities provided by ocean color satellite missions during the past decade have resulted in major advances in our ability to derive in-water biogeochemical properties linked to key processes within ocean ecosystems. The focus has primarily been on science in service of societal needs. However, ocean color has matured to the level where operational communities are looking to use the data directly, but there still needs to be a link to research that will improve our understanding of oceanic processes and undertake climatic research, e.g., the production of Essential Climate Variables (ECVs). As an example response to the need for ECV datasets, ESA set up the Climate Change Initiative programme that includes the Ocean Colour CCI project.

Assimilating biogeochemical properties, and in future IOPs, into ocean models is the focus of both research and operational activities, e.g., predicting CO_2 fluxes (Hemmings et al., 2008) and the wider remit GMES marine core service project called MyOcean/MyOcean2;

further details are available within “[Ocean Modeling and Data Assimilation](#).” Commercial operators are also realizing the value of ocean color-derived information, and its utilization among private sector customers is expected to increase significantly over the coming years due to a greater demand for marine environment information and high sensitivity to the cost-effectiveness of the different data-gathering approaches (IOCCG, 2008).

Bibliography

- Alvain, S., Moulin, C., Dandonneau, Y., and Breon, F. M., 2005. Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep-Sea Research I*, **1**, 1989–2004.
- Alvain, S., Moulin, C., Dandonneau, Y., Loisel, H., 2008. Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: a satellite view. *Global Biogeochemical Cycles*, **22**, doi:10.1029/2007GB003154.
- Antoine, D., and Morel, A., 1999. A multiple scattering algorithm for atmospheric correction of remotely sensed ocean colour (MERIS instrument): principle and implementation for atmospheres carrying various aerosols including absorbing ones. *International Journal of Remote Sensing*, **20**(9), 1875–1916.
- Arnone, R. A., Martinolich, P., Gould, R. W., Jr., Stumpf, R., and Ladner, S., 1998. Coastal optical properties using SeaWiFS. In *Proceedings Ocean Optics XIV*, Kailua Kona, HI, November 10–13, 1998.
- Badgley, P. C., and Childs, L. F., 1969. Earth resources surveys from space. In Badgley, P. C., Miloy, L., and Childs, L. (eds.), *Oceans from Space*. Houston, TX: Gulf, pp. 1–28.
- Barale, V., Larkin, D., Fusco, L., Melinotte, J. M., and Pittella, G., 1999. OCEAN Project: the European archive of CZCS historical data. *International Journal of Remote Sensing*, **20**(70), 1201–1218.
- Barnsley, M., Settle, J., Cutter, M., Lobb, D., and Teston, F., 2004. The PROBA/CHRIS mission: a low-cost smallsat for hyperspectral, multi-angle, observations of the Earth surface and atmosphere. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 1512–1520.
- Behrenfeld, M. J., and Falkowski, P. G., 1997. A consumer’s guide to phytoplankton primary productivity models. *Limnology and Oceanography*, **42**(7), 1479–1491.
- Behrenfeld, M. J., Westberry, T. K., Boss, E. S., O’Malley, R. T., Siegel, D. A., Wiggert, J. D., Franz, B. A., McClain, C. R., Feldman, G. C., Doney, S. C., Moore, J. K., Dall’Omo, G., Milligan, A. J., Lima, I., and Mahowald, N., 2009. Satellite-detected fluorescence reveals global physiology of ocean phytoplankton. *Biogeosciences*, **6**, 779–794.
- Berry, P. A. M., Smith, R. G., Salloway, M. K. and Benveniste, J., 2011. Global Analysis of EnviSat Burst Echoes Over Inland Water. *IEEE Transactions on Geoscience and Remote Sensing*, **99**, doi:10.1109/TGRS.2011.2170695.
- Campbell, J., Antoine, D., Armstrong, R., Arrigo, K., Balch, W., Barber, R., Behrenfeld, M., Bidigare, R., Bishop, J., Carr, M. E., Esaias, W., Falkowski, P., Hoepffner, N., Iverson, R., Kiefer, D., Lohrenz, S., Marra, J., Morel, A., Ryan, J., Vedernikov, V., Waters, K., Yentsch, C., and Yoder, J., 2002. Comparison of algorithms for estimating ocean primary production from surface chlorophyll, temperature, and irradiance. *Global Biogeochemical Cycles*, doi:10.1029/2001GB001444.
- Carder, K. L., Chen, F. R., Lee, Z. P., Hawes, S. K., and Cannizzaro J. P., 2003. *MODIS ocean science team algorithm theoretical basis document*. ATBD 19, Case 2 Chlorophyll-a. Version 7. 30 January 2003. College of Marine Science, University of South Florida. http://modis.gsfc.nasa.gov/data/atbd/atbd_mod19.pdf.
- Carr, M.-E., 2001. Estimation of potential productivity in Eastern Boundary Currents using remote sensing. *Deep-Sea Research II*, **49**, 59–80.
- Carr, M.-E., Friedrichs, M. A. M., Schmeltz, M., Aita, M. N., Antoine, D., Arrigo, K. R., Asanuma, I., Aumont, O., Barber, R., Behrenfeld, M., Bidigare, R., Buitenhuis, E. T., Campbell, J., Ciotti, A., Dierssen, H., Dowell, M., Dunne, J., Esaias, W., Gentili, B., Gregg, W., Groom, S., Hoepffner, N., Ishizaka, J., Kameda, T., Le Quere, C., Lohrenz, S., Marra, J., Melin, F., Moore, K., Morel, A., Reddy, T. E., Ryan, J., Scardi, M., Smyth, T., Turpie, K., Tilstone, G., Waters, K., and Yamanaka, Y., 2006. A comparison of global estimates of marine primary production from ocean color. *Deep-Sea Research II*, **53**, 741–770.
- Clarke, G. L., Ewing, G. C., and Lorenzen, C. J., 1970. Spectra of backscattered light from the sea obtained from aircraft as a measure of chlorophyll concentration. *Science*, **167**, 1119–1121.
- Cutter, M. A., Johns, L. S., Lobb, D. R., Williams, T. L., and Settle, J. J., (2003). Flight experience of the compact high resolution imaging spectrometer (CHRIS). In *Proceeding of SPIE Conference 5159 Imaging Spectrometry IX*, San Diego, CA, August 2003.
- Dierssen, H. M., Kudela, R. M., Ryan, J. P., and Zimmerman, R. C., 2006. Red and black tides: quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediments. *Limnology and Oceanography*, **51**(6), 2646–2659.
- Doerffer, R. and Schiller, H., 2000. Pigment index, sediment and gelbstoff retrieval from directional water leaving reflectances using inverse modelling technique, MERIS ATBD 2.12, Issue 4, Revision 0. <http://envisat.esa.int/instruments/meris/pdf/>.
- Doxaran, D., Froidefond, J.-M., Lavender, S., and Castaing, P., 2002. Spectral signature of highly turbid waters: application with SPOT data to quantify suspended particulate matter concentrations. *Remote Sensing of Environment*, **81**, 149–161.
- Frapparta, F., Calmanta, S., Cauhopé, M., Seyler, F., and Cazenave, A., 2006. Preliminary results of ENVISAT RA-2-derived water levels validation over the Amazon basin. *Remote Sensing of Environment*, **100**(2), 252–264.
- Friedrichs, M. A. M., Carr, M. E., Barber, R. T., Scardi, M., Antoine, D., Armstrong, R. A., Asanuma, I., Behrenfeld, M. J., Buitenhuis, E. T., Chai, F., Christian, J. R., Ciotti, A. M., Doney, S. C., Dowell, M., Dunne, J., Gentili, B., Gregg, W., Hoepffner, N., Ishizaka, J., Kameda, T., Limam, I., Marra, J., Mélin, F., Moore, J. K., Morel, A., O’Malley, R. T., O’Reilly, J., Saba, V. S., Schmeltz, M., Smyth, T. J., Tjiputraw, J., Waters, K., Westberry, T. K., and Winguth, A., 2009. Assessing the uncertainties of model estimates of primary productivity in the tropical Pacific Ocean. *Journal of Marine Systems*, **76**, 113–133.
- Frouin, R., Langer, D. W., Gautier, C., Baker, K., and Smith, R. C., 1989. A simple analytical formula to compute clear sky total and photosynthetically available solar irradiance at the ocean surface. *Journal of Geophysical Research*, **94**, 9731–9742.
- Frouin, R., Franz, B., and Wang, M., 2005. *Algorithm to estimate PAR from SeaWiFS data*. Version 1.2-documentation. http://oceancolor.gsfc.nasa.gov/DOCS/seawifs_par_wfifs.pdf.
- Garver, S. A., and Siegel, D. A., 1997. Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation: I. Time series from the sargasso sea. *Journal of Geophysical Research*, **102**, 18607–18625.
- Gordon, H. R., and Wang, M., 1994. Retrieval of water-leaving radiances and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. *Applied Optics*, **33**, 443–452.
- Gordon, H. R., Brown, O. B., and Jacobs, M. M., 1975. Computed relations between inherent and apparent optical properties of a flat homogeneous ocean. *Applied Optics*, **14**, 417–427.

- Gordon, H. R., Clark, D. K., Brown, J. W., Brown, O. B., Evans, R. H., and Broenkow, W. W., 1983. Phytoplankton pigment concentrations in the Middle Atlantic Bight: comparison of ship determinations and CZCS estimates. *Applied Optics*, **22**(1), 20–36.
- Gordon, H. R., Brown, J. W., and Evans, R. H., 1988. Exact Rayleigh scattering calculations for use with the Nimbus-7 Coastal Zone Scanner. *Applied Optics*, **27**(5), 862–871.
- Gower, J. F. R., and Borstad, G. A., 2004. On the potential of MODIS and MERIS for imaging chlorophyll fluorescence from space. *International Journal of Remote Sensing*, **25**(7), 1459–1464.
- Gower, J., Hu, C. M., Borstad, G., and King, S., 2006. Ocean color satellites show extensive lines of floating *Sargassum* in the Gulf of Mexico. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(12), 3619–3625.
- Gregg, W. W., Esaias, W. E., Feldman, G. C., Frouin, R., Hooker, S. B., McClain, C. R., and Woodward, R. H., 1998. Coverage opportunities for global ocean color in a multimission era. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 1620–1627.
- Gregg, W. W., Conkright, M. E., O'Reilly, J. E., Patt, F. S., Wang, M. H., Yoder, J. A., and Casey, N. W., 2002. NOAA-NASA coastal zone color scanner reanalysis effort. *Applied Optics*, **41**, 1615–1628.
- Hemmings, J. C. P., Barciela, R. M., and Bell, M. J., 2008. Ocean color data assimilation with material conservation for improving model estimates of air-sea CO₂ flux. *Journal of Marine Research*, **66**, 87–126.
- Holligan, P. M., Aarup, T., and Groom, S., 1989. The North Sea satellite colour atlas. *Continental Shelf Research*, **9**(8), 665–765.
- Hudson, H., 1608. A second voyage of Henry Hudson for finding a passage to the East Indies, by the north-east. In *Collections of the New-York Historical Society, for the Year, 1809*. New-York: Riley, Vol. 1, pp. 81–102.
- IOCCG, 1999. Status and plans for satellite ocean-colour missions: considerations for complementary missions. In Yoder, J. A. (ed.), *Reports of the International Ocean-Colour Coordinating Group*. Dartmouth: IOCCG, Vol. 2.
- IOCCG, 2006. Remote sensing of inherent optical properties: fundamentals, tests of algorithms, and applications. In Lee, Z. P. (ed.), *Reports of the International Ocean-Colour Coordinating Group*. Dartmouth: IOCCG, Vol. 5.
- IOCCG, 2008. Why ocean colour? The societal benefits of ocean-colour technology. In Platt, T., Hoepffner, N., Stuart, V., and Brown, C. (eds.), *Reports of the International Ocean-Colour Coordinating Group*. Dartmouth: IOCCG, Vol. 7.
- Kloiber, S. M., Brezonik, P. L., Olmanson, L. G., and Bauer, M. E., 2002. A procedure for regional lake water clarity assessment using Landsat multispectral data. *Remote Sensing of Environment*, **82**, 38–47.
- Korb, R. E., Whitehouse, M. J., and Ward, P., 2004. SeaWiFS in the southern ocean: spatial and temporal variability in phytoplankton biomass around South Georgia. *Deep-Sea Research II*, **51**, 99–116.
- Lavender, S., Doxaran, D., and Cherukuru, R. C. N., 2005. High spatial resolution remote sensing of the Plymouth coastal waters. In *3rd ESA CHRIS-PROBA Workshop CDROM*, 4 pp, ISBN 92-9092-904-9.
- Lee, Z. P., Carder, K. L., and Arnone, R., 2002. Deriving inherent optical properties from water color: a multi-band quasi-analytical algorithm for optically deep waters. *Applied Optics*, **41**, 5755–5772.
- Li, X., Pichel, W., Clemente-Colon, P., Krasnopolsky, V., and Sapper, J., 2001. Validation of coastal sea and lake surface temperature measurements derived from NOAA/AVHRR data. *International Journal of Remote Sensing*, **22**, 1285–1303.
- Lubin, D., Li, W., Dustan, P., Mazel, C. H., and Stamnes, K., 2001. Spectral signatures of coral reefs: features from space. *Remote Sensing of Environment*, **75**, 127–137.
- Maritorena, S., and Siegel, D. A., 2005. Consistent merging of satellite ocean color data sets using a bio-optical model. *Remote Sensing of Environment*, **94**, 429–440.
- Maritorena, S., Siegel, D. A., and Peterson, A. R., 2002. Optimization of semi-analytical ocean color model for global scale applications. *Applied Optics*, **41**, 2705–2714.
- MERIS, 2007. Algorithm theoretical basis document 2.9: pigment index retrieval in Case 1 waters. <http://envisat.esa.int/instruments/meris/pdf/>.
- Miller, R. L., and McKee, B. A., 2004. Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters. *Remote Sensing of Environment*, **93**, 259–266.
- Moore, G. F., Aiken, J., and Lavender, S. J., 1999. The atmospheric correction of water colour and the quantitative retrieval of suspended particulate matter in Case II waters: application to MERIS. *International Journal of Remote Sensing*, **20**(9), 1713–1733.
- Morel, A., 1974. Optical properties of pure seawater. In Jerlov, N. G., and Nielsen, S. E. (eds.), *Optical Aspects of Oceanography*. London: Academic.
- Morel, A., and Gentili, B., 1993. Diffuse reflectance of oceanic waters. II. Bidirectional aspects. *Applied Optics*, **32**(33), 6864–6879.
- Morel, A., and Maritorena, S., 2001. Bio-optical properties of oceanic waters: a reappraisal. *Journal of Geophysical Research*, **106**, 7163–7180.
- Morel, A., and Prieur, L., 1977. Analysis of variations in ocean color. *Limnology and Oceanography*, **22**, 709–722.
- Mueller, J. L., Bidigare, R. R., Trees, C., Balch, W. M., Dore, J., Drapeau, D. T., Karl, D., Van Heukelem, L., and Perl, J., 2003. Ocean optics protocols for satellite ocean color sensor validation. In Mueller, J. L., Fargion, G. S., and McClain, C. R. (eds.), *Biogeochemical and Bio-Optical Measurements and Data Analysis Protocols*, 5th edn. Greenbelt, MD: NASA Goddard Space Flight Center. NASA Technical Memorandum 2003, Vol. V, pp. 1–35.
- Mumby, P. J., and Edwards, A. J., 2003. Mapping marine environments with IKONOS imagery: enhanced spatial resolution can deliver greater thematic accuracy. *Remote Sensing of Environment*, **82**(2–3), 248–257.
- Mumby, P. J., Chisholm, J. R. M., Clark, C. D., Hedley, J. D., and Joubert, J., 2001. A bird's-eye view of the health of coral reefs. *Nature*, **413**, 36.
- Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M. H., Devred, E., and Bouman, H., 2008. Remote sensing of phytoplankton functional types. *Remote Sensing of Environment*, **112**, 3366–3375.
- Neukermans, G., Ruddick, K., Bernard, E., Ramon, D., Nechad, B., and Deschamps, P.-Y., 2009. Mapping total suspended matter from geostationary satellites: a feasibility study with SEVIRI in the Southern North Sea. *Optics Express*, **17**, 14029–14052.
- O'Reilly, J. E., Maritorena, S., Mitchell, B. G., Siegel, D. A., Carder, K. L., Garver, S. A., Kahru, M., and McClain, C., 1998. Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, **103**(C11), 24937–24953.
- O'Reilly, J. E., et al., 2000. *SeaWiFS Postlaunch Calibration and Validation Analyses*. Greenbelt, MD: NASA Goddard Space Flight Center. SeaWiFS Postlaunch Technical Report Series, Vol. 11, Part 3.
- Oesch, D. C., Jaquet, J. -M., Hauser, A., and Wunderle, S., 2005. Lake surface water temperature retrieval using advanced very high resolution and Moderate Resolution Imaging Spectroradiometer data: validation and feasibility study. *Journal of Geophysical Research*, **110**(C12014), doi:10.1029/2004JC002857.

- Pinkerton, M. H., Moore, G. F., Lavender, S. J., Gall, M. P., Oubelkheir, K., Richardson, K. M., Boyd, P. W., and Aiken, J., 2006. A method for estimating inherent optical properties of New Zealand continental shelf waters from satellite ocean colour measurements. *New Zealand Journal of Marine and Freshwater Research*, **40**, 227–247.
- Pinnock, S., Fanton Anton, O., and Lavender, S., 2008. GlobColour: a precursor to the GMES Marine Core Service Ocean Colour Thematic Assembly Centre. *ESA Bulletin*, **132**, 43–49.
- Platt, T., 1986. Primary production of the ocean water column as a function of surface light intensity: algorithms for remote sensing. *Deep Sea Research*, **33**, 149–163.
- Platt, T., and Sathyendranath, S., 2008. Ecological indicators for the pelagic zone of the ocean from remote sensing. *Remote Sensing of Environment*, **112**, 3426–3436.
- Pradhan, Y., Lavender, S. J., Hardman-Mountford, N. J., and Aiken, J., 2006. Seasonal and inter-annual variability of chlorophyll-a concentration in the Mauritanian upwelling: observation of an anomalous event during 1998–1999. *AMT Deep-Sea Research II Special Issue*, **53**, 1548–1559.
- Raitsos, D. E., Lavender, S. J., Maravelias, C. D., Haralambous, J., Richardson, A. J., and Reid, P. C., 2008. Identifying four phytoplankton functional types from space: an ecological approach. *Limnology and Oceanography*, **53**, 605–613.
- Raitsos, D. E., Lavender, S. J., Pradhan, Y., Tyrrell, T., Reid, P. M. C., and Edwards, M., 2006. Coccolithophore bloom size variation in response to the regional environment of the subarctic North Atlantic. *Limnology and Oceanography*, **51**(5), 2122–2130.
- Rast, M., Bezy, J. L., and Bruzzi, S., 1999. The ESA medium resolution imaging spectrometer MERIS – a review of the instrument and its mission. *International Journal of Remote Sensing*, **20**(9), 1679–1680.
- Ruddick, K. G., Ovidio, F., and Rijkeboer, M., 2000. Atmospheric correction of SeaWiFS imagery for turbid and inland waters. *Applied Optics*, **39**, 897–912.
- Sathyendranath, S., Watts, L., Devred, E., Platt, T., Caverhill, C. M., and Maass, H., 2004. Discrimination of diatoms for other phytoplankton using ocean-colour data. *Marine Ecology Progress Series*, **272**, 59–68.
- Schiller, H., and Doerffer, R., 1999. Neural network for emulation of an inverse model – operational derivation of Case II water properties from MERIS data. *International Journal of Remote Sensing*, **20**(9), 1735–1746.
- Shutler, J. D., Land, P. E., Smyth, T. J., and Groom, S. B., 2007. Extending the MODIS 1 km ocean color atmospheric correction to the MODIS 500 m bands and 500 m chlorophyll-a estimation towards coastal and estuarine monitoring. *Remote Sensing of Environment*, **107**, 521–532.
- Siegel, D. A., Wang, M., Maritorena, S., and Robinson, W., 2000. Atmospheric correction of satellite ocean color imagery: the black pixel assumption. *Applied Optics*, **39**, 3582–3591.
- Simpson, J. J., 1994. Remote sensing in fisheries: a tool for better management in the utilization of a renewable resource. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 743–771.
- Smyth, T. J., Moore, G. F., Hirata, T., and Aiken, J., 2006. Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment. *Applied Optics*, **45**(31), 8116–8131.
- Sterckx, S., Debruyne, W., Vanderstraete, T., Goossens, R., and Van der Heijden, P., 2005. Hyperspectral data for coral reef monitoring. A case study: Fordate, Tanimbar, Indonesia. *EARSeL eProceedings*, **4**(1), 18–25.
- Stumpf, R. P., Arnone, R. A., Gould, R. W., Martinolich, P., and Ransibrahmanakul, V., 2002. A partially-coupled ocean–atmosphere model for retrieval of water-leaving radiance from SeaWiFS in coastal waters. In Hooker, S. B., and Firestone, E. R. (eds.), *Algorithm Updates for the Fourth SeaWiFS Data Reprocessing*. Greenbelt, MD: NASA Goddard Space Flight Center. NASA Technical Memorandum 2002–206892, Vol. 22.
- Tester, P. A., and Stumpf, R. P., 1998. Phytoplankton blooms and remote sensing: what is the potential for early warning. *Journal of Shellfish Research*, **17**(5), 1469–1471.
- Van Mol, B., and Ruddick, K., 2005. Total suspended matter maps from CHRIS imagery of a small inland water body in Oostende (Belgium). In *Proceedings of the 3rd ESA CHRIS/Proba Workshop*, March 21–23, 2005, Frascati: ESRINI, (ESA SP-593, June 2005).
- Vincent, R. K., Qin, X., McKay, R. M. L., Miner, J., Czajkowski, K., Savino, J., and Bridgeman, T., 2004. Phycocyanin detection from Landsat TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sensing of Environment*, **89**, 381–392.
- Wang, M., and Shi, W., 2005. Estimation of ocean contribution at the MODIS near-infrared wavelengths along the east coast of the U.S.: two case studies. *Geophysical Research Letters*, **32**, L1360.
- Wernand, M., 2008. The Arcane colouring of natural waters; From Hudson (1600) to Raman (1930). *NIOZ Internal Report, Physics*, **1**, 128.
- Wilson C., and Adamec, D., 2002. A global view of biophysical coupling from SeaWiFS and TOPEX satellite data, 1997–2001. *Geophysical Research Letters*, **29**, 10.1029/2001GL014063.
- Zimmermann, G., and Neumann, A., 1997. MOS a spaceborne imaging spectrometer for ocean remote sensing. *Backscatter*, **8**, 2. <http://www.iocccg.org/reports/mos/mos.html>.

Cross-references

Fisheries
 Ocean Modeling and Data Assimilation
 Ocean Surface Topography
 Sea Surface Temperature
 Sea Surface Wind/Stress Vector

OCEAN MODELING AND DATA ASSIMILATION

Detlef Stammer
 Institut für Meereskunde, Zentrum für Marine und
 Atmosphärische Wissenschaften, Universität Hamburg,
 Hamburg, Germany

Synonyms

Numerical simulations; Ocean state estimate; Reanalysis

Definition

Ocean modeling. Finding solutions to the equations of motion describing the movement of the ocean fluid and its properties as a function of time and geographic location.

Data assimilation. Constraining ocean model solutions by observations to obtain a more realistic description of the ocean.

Ocean modeling

Introduction

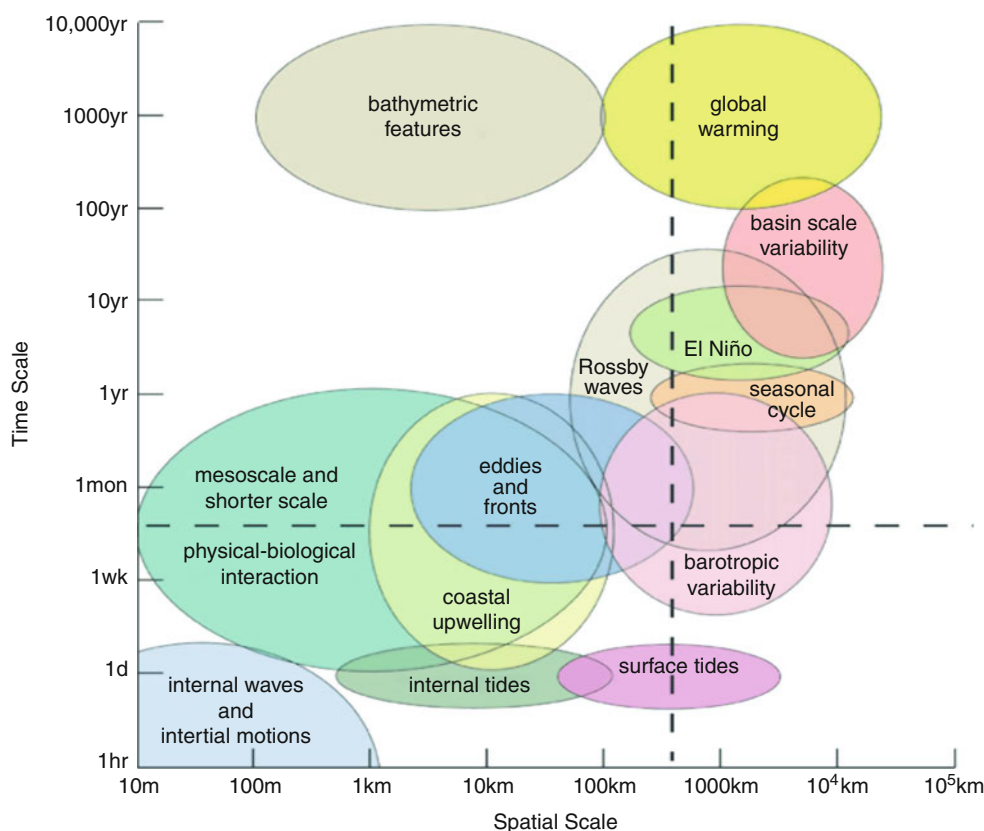
Ocean modeling, like atmospheric modeling and numerical weather forecast, is a mature field of computational fluid dynamics. It is concerned with a dynamical description of the general ocean circulation, aiming at (1) simulations of the time-varying ocean circulation or aspects of it, (2) providing a dynamical description of the time-varying ocean circulation, (3) simulating its transport properties, and (4) quantifying the relation between the movement of water, forces acting at the ocean's boundaries, and dissipation and mixing processes retarding the flow. As tools, ocean modeling uses *numerical circulation models* to solve the dynamical *equations of motion* that govern the motion of the ocean's flow in response to forces acting on it. An overview to ocean modeling is provided by, e.g., Haidvogel and Beckmann (1999).

The ocean is a forced-dissipative fluid, with forcing essentially provided at the boundaries and dissipation taking place everywhere at the molecular scale. The volume of seawater is bounded by complex land-sea boundaries and motions of seawater are constrained by rotation of the planet and stratification of the water column. In response to forces acting especially at the sea surface of

the ocean, flow in the ocean interior exhibits a rich mix of dynamical phenomena, including boundary currents, large-scale gyres and jets, boundary layers, linear and nonlinear waves, and quasi-geostrophic and three-dimensional turbulence. Those phenomena are summarized in Figure 1 along with their spatial and temporal scales.

Clearly, motions of the ocean cover temporal scales ranging from less than an hour (e.g., surface waves) to hundreds of years (deep circulation and mixing processes) and spatial scales ranging from a few millimeters (turbulence and diffusion) to the circumference of the Earth (global circulation). An overview of processes acting as part of the ocean flow field is provided by Steward (2005) and dynamics are described, e.g., by Vallis (2006).

Models are being used experimentally to investigate hypotheses for ocean phenomena, to simulate and nowcast the observed flow, to consider future scenarios such as those associated with human-induced climate warming, and to forecast ocean conditions on weekly to decadal timescales using dynamical modeling systems. In the context of remote sensing, numerical models of the ocean became an important tool for interpreting and correcting *satellite observations of the ocean* (see below).



Ocean Modeling and Data Assimilation, Figure 1 Dynamical phenomena occurring in the ocean and their space and timescales plotted against their respective space and timescales (D. Chelton, pers. communication.).

Equations of motion

The equations of motion, governing the circulation of the ocean and the atmosphere alike, are based on Newtonian mechanics and irreversible thermodynamics applied to a continuum fluid. The conservation of momentum builds the backbone for the dynamical description of the flow providing information of the three-dimensional flow field. Together with additional conservation equations of heat and material constituents, such as potential temperature or salt, they comprise a system of scalar equations, called the Navier–Stokes equations. These equations exist in various levels of approximation. In compressible form and z -level representation, the equations can be expressed as:

$$\begin{aligned} \frac{D\mathbf{v}_h}{Dt} + f\mathbf{k} \times \mathbf{v}_h + \frac{1}{\rho} \nabla_z p &= \mathbf{F} \\ \frac{Dw}{Dt} + \frac{gp}{\rho_0} + \frac{1}{\rho_0} \frac{\partial p}{\partial z} &= F_w \\ \frac{1}{\rho_0 c_s^2} \frac{Dp}{Dt} + \nabla_z w + \frac{\partial w}{\partial z} &= 0 \\ \rho &= \rho(\theta, S, p) \\ \frac{D\theta}{Dt} &= Q_\theta \\ \frac{DS}{Dt} &= Q_s \end{aligned}$$

Here $D/Dt = \partial/\partial t + \mathbf{u} \cdot \nabla$ is the total derivative, $\nabla = \partial/\partial x + \partial/\partial y + \partial/\partial z$ is the nabla operator, and $\nabla_z = \partial/\partial z$ is the vertical component of the nabla operator. $\mathbf{v}_h = (\mathbf{u}, \mathbf{v})$ is the two-dimensional horizontal flow field; w is the vertical component of the flow field; ρ is the density of sea water with ρ_0 being a reference (mean) density; p is the hydrostatic pressure; $\mathbf{F}$ represents forcing and dissipation acting on momentum; Q_θ and Q_s are sources of potential temperature and salt; and c_s is the speed of sound.

Usually the equations of motion as stated above or in other forms of approximation are too difficult to solve analytically, largely due to the nonlinear and turbulent nature of the ocean's flow, involving processes from the mm scale (e.g., dissipation of energy) up to basin and global scale spanning thousands of km, and the very long time-scales (decades to centuries). Over those scales, water mass properties, e.g., potential temperature, salinity, oxygen, CO_2 , or nutrients, are preserved in the ocean interior and ocean models need to show similar conservation properties. All those difficulties together require the use of computers to solve the equations numerically and to use numerical models to explore the immense phase space of solutions.

A detailed account for approximations of the Navier–Stokes equations and the effect of the approximations on the dynamical description of the ocean is provided by Müller (2006) and Olbers et al. (2012). As an example, there are two approximations that independently

filter out acoustic modes. The *non-divergence approximation* (associated with *Boussinesq fluids*) removes three-dimensional acoustic waves, and the hydrostatic balance removes vertical acoustic waves.

Numerical implementation

Numerical solutions of the equations of motion (or any other differential equation) are obtained by approximating differential operators in space and time through difference operators involving finite length or finite volume elements of the model domain:

$$\begin{aligned} \frac{dx}{dt} &\approx \frac{\Delta x}{\Delta t} = \frac{x_i - x_{i-1}}{t_i - t_{i-1}}; \\ \frac{d^2x}{dt^2} &\approx \frac{\Delta(\Delta x)}{(\Delta t)^2} = \frac{x_{i+1} - 2x_i + x_{i-1}}{(t_i - t_{i-1})^2} \end{aligned}$$

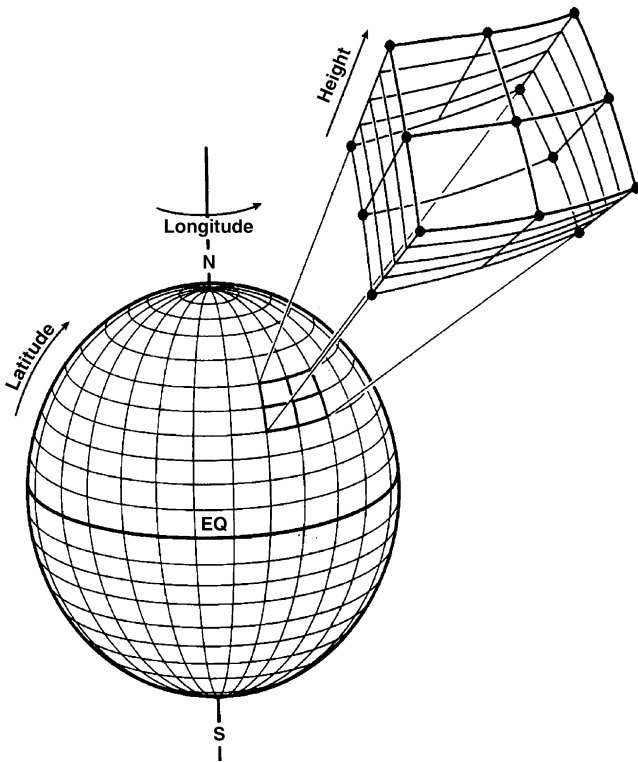
Here t_i can be a time index or a space index. Written in this way, the model domain is split into many elements, each of which represents a small fluid parcels (Figure 2). Each variable then represents a mean value of the finite element of the model domain, typically reaching between 10 and 100 km horizontally.

Rewriting the equations of motion in a finite difference form allows for computing the model variables (also called state variables) for each time step from those available during the last time step accounting for forcing terms. Applied many times, this process leads to an initial value problem and boundary value problem for the state vector, $\mathbf{x}(t)$:

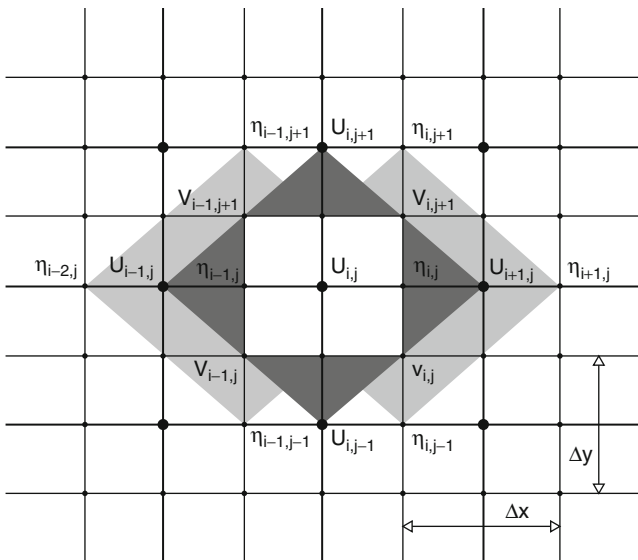
$$\mathbf{x}(t + 1, i, j, k) = \mathbf{A}(t)\mathbf{x}(t, i, j, k) + \mathbf{F}(t)$$

The state vector, $\mathbf{x}$, which exists on the typically regular model grid, includes values of temperature, salinity, pressure, and velocity. Here $\mathbf{A}$ is the step operator stepping the model forward from one time step to the next, and in this form one needs to be provided initial conditions from which the equation is being integrated forward, surface forcing terms ($\mathbf{F}(t) = \mathbf{G}(t)\mathbf{u}(t)$, including wind stress, heat, and freshwater fluxes as well as river runoff) and subgrid-scale parameterization. Also, information about the bottom topography (shape of the sea floor) and coastlines needs to be provided. From the state vector, $\mathbf{x}$, one can compute any derived quantity (e.g., enstrophy, potential vorticity, or enthalpy flux) of interest.

There are various ways to arrange the full suite of model variables horizontally, usually referred to as Arakawa staggered grids (Haidvogel and Beckmann, 1999). An example of such a grid is given in Figure 3, representing a so-called Arakawa C-grid. On this grid, the velocity variables and temperature and salinity are shifted relative to each other in space. As an example, the MITgcm is specified on an Arakawa-C grid (e.g., Marshall et al., 1997). Discretization is also applied in the vertical, dividing the ocean in layers, e.g., of constant depth or constant density.



Ocean Modeling and Data Assimilation, Figure 2 Horizontal model grid (<http://acmg.seas.harvard.edu/curresh.html>).



Ocean Modeling and Data Assimilation, Figure 3 Example of a horizontal model grid, illustrating an Arakawa-C grid.

Ocean models are typically characterized in terms of geography, physics, surface approximations, vertical discretization, and density variation. Treating the ocean in finite elements also requires a description of processes

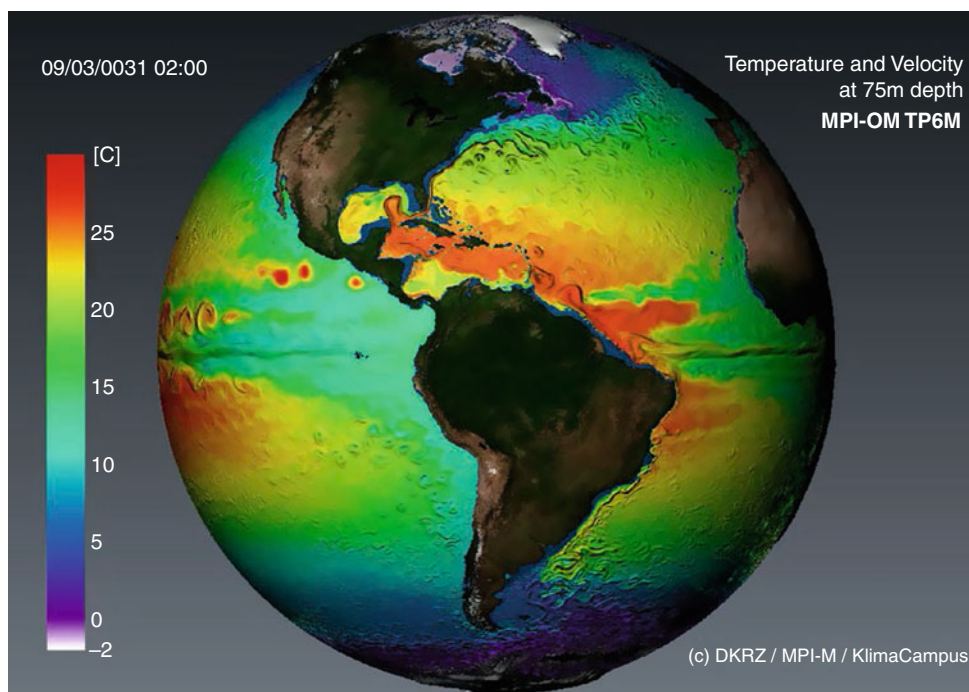
that are happening on subgrid scales in terms of resolved fields (*turbulence closure*). Griffies and Adcroft (2008) discuss numerous questions that arise when discretizing the ocean equations, such as how to conserve properties tracer conservation and consistency with mass conservation. Ocean codes that fail to respect these properties are severely handicapped for use in studies of ocean dynamics or climate.

Applications

The present status of ocean modeling and anticipated improvements is addressed by Griffies et al. (2001, 2010). Applications of ocean models for studies of fluid dynamics in general include direct numerical simulations (DNS) where attempts are being made to resolve all processes, large-scale eddy simulations (“LES”) where subgrid-scale processes are being presented by the properties of the large-scale flow, and long, coarse-resolution climate simulations. The first class of simulations is much akin to what is being performed in computational fluid dynamics (CFD), where an attempt is made to directly solve all relevant phenomena without involving subgrid-scale parameterizations. Ocean climate modeling represents the other end of the spectrum and is an application of a very different nature to those found in other areas of CFD. Here the timescales of interest are decades to millennia, yet simulations require resolution or parameterization of phenomena whose timescales are minutes to hours. Furthermore, the most energetic spatial scales are of order 10–100 km (mesoscale eddies), yet the problem is fundamentally global in nature. Ocean modeling can also be subdivided into classes of applications:

1. Process-oriented modeling
2. Operational now- and forecasting
3. Climate-oriented sensitivity experiments of predictions

An example of state-of-the-art eddy-resolving but climate-oriented ocean modeling is given in Figure 4, showing the simulation of the near-surface flow field of the ocean. The field is taken from the German consortium effort STORM (www.dkrz.de/STORM; see von Storch et al., 2012 for details) and was run at the German Climate Computer Center (DKRZ). The goal of the project is to run a coupled climate model in eddy-resolving mode parallel to traditional lower-resolution versions of the same model used as part of the IPCC AR5 process and thereby to investigate the impact of “weather” in the atmosphere and ocean and the projection of climate variability and climate change. Shown in the figure is an instantaneous field of the speed of the ocean in about 75 m depth. The figure clearly reveals the chaotic, turbulent nature of the flow field. Ocean modeling aims to improve our understanding of the underlying physical processes, the interaction of the turbulent flow field with biological and geochemical processes in the ocean, including primary production and CO₂ uptake.



Ocean Modeling and Data Assimilation, Figure 4 Near-surface flow and temperature fields as the results from the German STORM effort in 75 m depths. Shown are the temperature in °C (color) and the speed (shading) at one instant. Both fields can be observed at the surface by satellites using radiometry and altimetry (see von Storch et al., 2012 for details).

Importance of satellite data for ocean modeling

In terms of numerical simulations of the ocean, satellite observations play a major role in testing and improving ocean models. This is being done through a comparison of models and observations (e.g., eddy variability) because satellite data provide a unique database of the time-varying features at the sea surface which is nearly impossible to observe with conventional technology. Satellite data also build a very important basis for improving models through data assimilation. At the same time ocean models (with and without data assimilation) are needed to interpret satellite data. A typical application in this respect is that of observing surface temperature using remote sensing. Because infrared radiometers on satellites measure the emission from only the top few tenths of a mm of the water column, which is typically affected by air-sea interaction, we need ocean models to extend the measurements into the water beneath. This includes the correction of the ocean thermal skin effect and the daily cycle in temperature; these also have to be taken into account before observations taken at different times of the day can be merged into a single field.

Data assimilation

Introduction

Data assimilation is a subfield of ocean modeling. Its goal is to increase the realism of model simulations by merging them with ocean data and by doing so to improve the

model state, model parameters, or both. This is akin to practices in atmospheric weather prediction. However, there are several important differences. In particular, weather prediction is targeted at best possible forecasts. In contrast, many applications of assimilation in the ocean are mostly targeted at describing the history of ocean variability and the present state of the ocean (so-called nowcasts): Ocean state estimation (data assimilation) has as its goal to obtain the best possible description of the changing ocean by combining, in some suitably optimum way, all the diverse ocean observations with theoretical knowledge of the ocean circulation as embodied in numerical ocean circulation models. If carried out properly, the result is a dynamically self-consistent estimate of the time-evolving ocean circulation, one, which has greater information and forecast skill than does either model or data alone. Aspirations of forecasting are to provide seasonal to interannual predictions. However, they are gaining increased attention now in the context of ocean and climate services, including decadal climate forecasts.

Assimilation approaches

Regardless for which purpose and with which detailed approaches pursued, state estimation or “data assimilation” in general is just least-squares fitting of models to data, taking into account the model equations as constraints. Regardless of the details, all approaches attempt to minimize a model-data misfit. In doing so, assimilation methods deal, in principle, with the state vector $\mathbf{x}(t)$,

representing again a time-dependent simulation of the state of the ocean. All oceanic measurements can be associated with its model equivalent through a functional relationship that can be written as:

$$\mathbf{E}(t)\mathbf{x}(t) + \mathbf{n}(t) = \mathbf{y}(t),$$

where $\mathbf{E}$ is a matrix, $\mathbf{n}(t)$ is the measurement error including data noise, and $\mathbf{y}(t)$ is the vector of observations. An estimated time-dependent solution, $\mathbf{x}(t)$, is sought that minimizes the model-data misfits, which then can be written as:

$$J = \sum_t (\mathbf{y}(t) - \mathbf{E}(t)\mathbf{x}(t))^T \mathbf{R}(t) (\mathbf{y}(t) - \mathbf{E}(t)\mathbf{x}(t)) + \mathbf{x}_0^T \mathbf{Q} \mathbf{x}_0 + \mathbf{u}^T \mathbf{P} \mathbf{u}$$

subject to the model dynamics as embedded in the numerical model code. $\mathbf{R}(t)$, $\mathbf{Q}$, and $\mathbf{P}$ are the error covariances representing observations, the model, and the model parameters. The choice of those three covariances determines the solution as they select one solution from the multitude of all possible solutions.

As described in detail by Wunsch (2006), there are many methods for solving *constrained least-squares problems*, either exactly, by iteration, or sequentially. In terms of nomenclature, they include Nudging, 4DVAR, 3DVAR, adjoint, objective interpolation (OI), Kalman filter (KF), RTS smoother, ensemble KF, Pontryagin principle, relaxation, line-searches, breeding vectors, SVD, and many more. All these apparently different methods are nonetheless just variant algorithms used to find the minimum of an objective (or cost) function measuring the deviation of the model's state from observations. They essentially differ in the extent to which an approximation to that minimum is acceptable, whether one intends to find a dynamically self-consistent solution and whether or not one seriously seeks an estimate of the error in the result.

Much of the assimilation technology used today in oceanography was first developed for atmospheric application. However, for several decades, ocean applications evolved as well and now constitute a sustained activity in support of ocean services and climate research. Underlying assimilation schemes range from simple and computationally efficient (e.g., optimal interpolation) to sophisticated and computationally intensive (e.g., adjoint, Kalman filters, and smoothers). However, it is also worth mentioning that as compared to atmospheric systems, some ocean estimation approaches are dynamically self-consistent, guaranteeing the conservation of heat, freshwater, and momentum. This is an important requirement for climate applications of reanalysis results, and here the ocean applications are ahead of the atmospheric applications (see also Bengtsson et al., 2007).

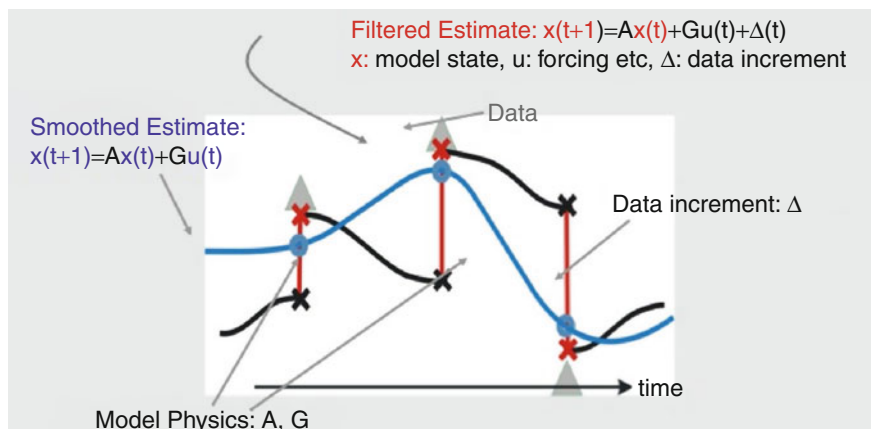
Differences between filters and smoothers are schematically illustrated in Figure 5. Filters lead to a new model state, once observations become available, by merging the simulated state with the observations, thereby

neglecting all model conservation equations. In contrast, smoothers attempt to find uncertain model parameters by using observations from the past and the future, which then lead to an optimal simulation of the observations by the improved model. Simplified methods (e.g., nudging, robust diagnostic, and objective mapping), thus, intended to find approximations to the minimum of J by relaxing the model constraints. They are easy to set up and are computationally inexpensive. However, the temporal evolution of such ad hoc solutions is not necessarily dynamically consistent and usually implies internal sinks and sources of momentum, heat, and freshwater. Rigorous methods, such as provided by an adjoint approach, are computationally much more demanding. We note, however, that they are needed to obtain dynamically self-consistent ocean estimates useful for understanding the physics of the system by exploiting all the information contained in the data.

Important milestones in the development of ocean state estimation included the development of *inverse methods* (adjoint, Kalman filter, etc.) that can be applied to ocean circulation models using supercomputers. As an example, the development of ocean adjoint models encompassed the existence of modern primitive equation (PE) models, the development of automatic differentiation tools (Giering and Kaminsky, 1998), their pilot applications to ocean problems (Marotzke et al., 1999; Stammer et al., 2002a), and for that purpose, enhancements and adjustments of the computer infrastructure to finally encompass long optimization runs. In almost all cases, achieving accomplishments in the development of the infrastructure required nearly a decade of sustained consortium efforts (Stammer et al., 2002b). This development has proven to be a large endeavor that requires expertise in ocean observations, modeling, assimilation, as well as information technology, which needs to be sustained and to have a long-term perspective to be effective. If ocean state estimation were to be established as a firm part of an ocean or climate information system, a sustained effort will be needed for the synthesis part.

Applications of ocean assimilation

Present applications of data assimilation in the ocean can be subdivided roughly into three application areas: (1) operational ocean oceanography and nowcasting, (2) climate-oriented data synthesis, and (3) initialization of predictions in the ocean or of coupled climate models. Similar to the applications, approaches also differ, ranging from simple but computational cheap nudging techniques to computational demanding but mathematically rigorous methods. A summary of most of those efforts can be found in Lee et al. (2010) and Reinicker et al. (2010). The high frequency and high spatial coverage of satellite data now allow for global high resolution operational ocean products produced by efforts such as MERCATOR, MyOcean, NAOCEANO, etc.). See Bell et al. (2009) for details.



Ocean Modeling and Data Assimilation, Figure 5 Schematic of data assimilation (From I. Fukumori, JPL, Personal communication).

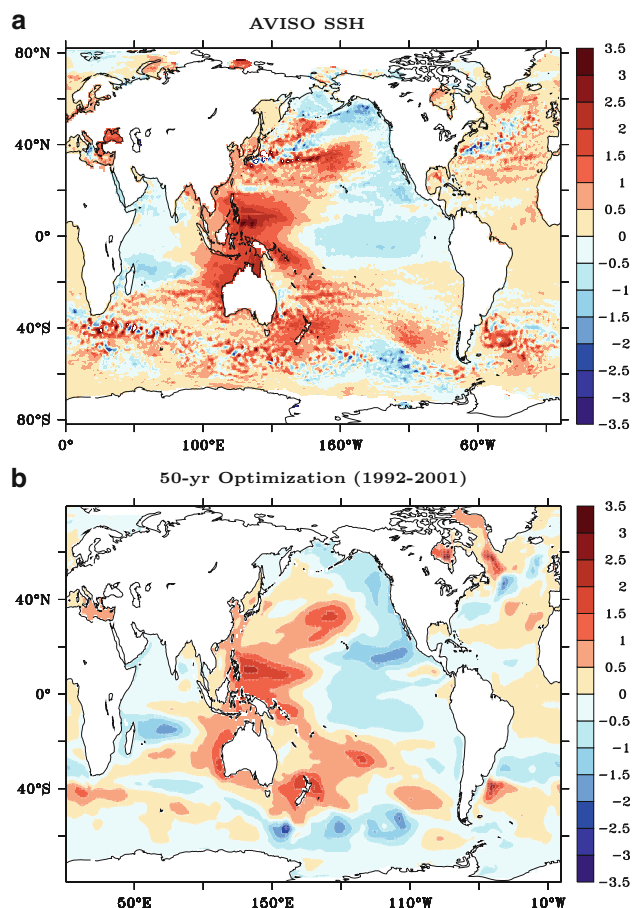
At the time of writing, several successful ocean syntheses exist which are being exploited in terms of ocean dynamics, their potential as initial conditions for climate models or simply to describe ocean variability over the last 50+ years. Ocean efforts are targeting long-duration descriptions of the ocean. Some of the existing assimilation products span the period of the past several decades and are relatively coarse in spatial resolution; others focus on the data rich period, roughly from 1992 to present and tend to be eddy permitting, and sometimes only regional in extent. Typical applications include the estimate of ocean transports, change in the oceans' heat and freshwater content, transports of heat and salt, the interaction of the ocean with the atmosphere in terms of heat and freshwater fluxes through the ocean surface, or the important aspect of sea level variability and change. In the context of the global energy balance, questions specific to the ocean include the following: What is the present rate of change of heat storage in the ocean over the last 10 years and the last 50 years and how is it distributed with depth and location? What are the key regions and their significance of ocean–atmosphere feedbacks? What is the meridional heat transport of the oceans and its variability on seasonal, interannual, and decadal timescales? What is the variability on global scales of the deep and shallow meridional overturning circulations and of the interoceanic exchange of heat? How does the global ocean heat storage relate to imbalance in the planetary radiation budget?

During recent years, several data-based estimates of changes of the ocean heat content and freshwater were published from which the emerging picture of the upper ocean's heat content changed substantially, partially due to correction of errors found in the underlying data sets (XBT and ARGO), and these improvements will continue in the future. In this context, dynamically consistent state estimation can make especially valuable contributions by providing estimates of climate-relevant indices or quantities and also by providing estimates of residuals in the data

not compatible with the dynamics and other components, such as surface forcing, thereby giving important information about potential data errors not yet accounted for (see also below).

A quantity of general concern is sea level and its variability, which represents an integral over many individual aspects of the ocean state. Questions related to [sea level rise](#) are among those most directly related to anthropogenic climate change and its impacts. Changes in sea level can have a potentially substantial impact on society and understanding of ongoing and past changes and their regional character is therefore of specific importance. [Figure 6](#) shows estimates of local sea-surface height (SSH) changes as they follow from observations and from the GECCO estimates (Köhl and Stammer, 2008). We realize that large-scale trends in the simulated SSH agree with observed trends over the last 15+ years of altimetric satellite observations. Over the large parts of the world's ocean, GECCO results suggest that SSH changes induced by heat content changes and changes in salinity counterbalance, to some extent, as this either would follow from advection of water masses along isopycnals (e.g., wind-driven changes in GECCO) or imposed by the use of T/S relations.

The focus of ocean data assimilation has now shifted more to aspects of forecasting the ocean and climate system in support of ocean services and climate services. Work has been continuing now for several years in the context of seasonal predictions of the climate system (Anderson, 2010; Balmaseda et al., 2010) and is now being extended towards seamless predictions in the ocean and of the coupled climate system. To perform predictions in the ocean, one needs proper initial conditions as well as surface forcing fields. While the latter have to come from an atmospheric or a coupled ocean–atmosphere forecast model, the initial conditions of the ocean have to be prescribed through suitable ocean measurements, including satellite observations of the ocean. Theoretically, initial conditions should represent accurately the state of the



Ocean Modeling and Data Assimilation, Figure 6 Trends in sea level as observed by the TOPEX/Poseidon altimeter mission (top) over the period 1993–2001 and as simulated by the GECCO 50 year ocean state estimation approach (Köhl and Stammer, 2008). Units are cm/year.

real world. However, given the fact that measurements have uncertainties, this would not necessarily be the best initial condition for optimizing the forecast skill. Among the practical ways of creating ocean initial conditions is to estimate the initial conditions of a coupled climate model. An application of ocean state estimation is that of ocean initialization for seasonal and decadal forecasting.

Importance of satellite data for ocean assimilation

The skill of data assimilation depends on the quality of underlying models, but also on the quantity and quality of available observations. While the amount of ocean data has been substantially increasing since the 1990s, satellite data are still, and will remain, the largest source of information obtained about the ocean quasi-globally and in quasi-real time. On the other hand, because the ocean is essentially opaque to electromagnetic radiation, satellite data by itself represent first of all the surface of the ocean (excepting the visible part of the electromagnetic

spectrum, where the penetration depth of light is up to about 50 m). In the context of remote sensing, ocean modeling and data assimilation therefore have the specific role of extrapolating information available at the ocean surface and connecting it to the three-dimensional ocean circulation. Thus, data assimilation is based on the understanding that a complete picture of the ocean for the purpose of climate research and applications will come only from a synergy between observations, modeling, and data assimilation.

As in atmospheric applications (e.g., weather forecasting), the quality of observed ocean data is a major concern for ocean state estimates, as are the changes in data coverage due to changes in observing systems, changes in quality of the data by improving instrumentation, or even new types of observations available since recently, e.g., time-varying gravity from US/German GRACE mission or salinity from the ESA SMOS satellite and the US-Argentine Aquarius/SAC-D mission. As an example, the lack of salinity data before the deployment of Argo floats severely hampers our ability to describe decadal variability of the ocean circulation because of the salinity influence on density and thus on the ocean circulation. Over recent decades, satellite data (e.g., altimeters and scatterometers) have proven to be of utmost importance in producing a global view of the ocean and especially in providing information about small space- and time-scales hardly available from other observing components. Much of the future progress hinges on continuity of these satellite missions and the ability to produce climate data records.

Summary

Ocean modeling and assimilation have both become a mature field within oceanography. *Ocean modeling* is concerned with finding solutions to the equations of motion describing the movement of the ocean fluid and its properties as a function of time and geographic location. Applications range from process studies of the ocean on almost all space and timescales, but include also simulations required to improve the analyses of satellite data. In contrast, *data assimilation* is concerned with constraining ocean model solutions by observations to obtain a more realistic description of the time-varying ocean circulation and its interaction with the atmosphere or sea ice. In both cases, satellite data play a major role, for improving models through model-data comparison studies, through merging the data with models through the approach of data assimilation or through the use of observations as initial conditions. The future will see many more applications of all of those aspects. It can be envisioned that model resolution will be increased further, making results more realistic and applicable, especially near frontal structures in the ocean or ocean boundaries. Satellite data will continue to play a key role in the process of data assimilation and increased focus will be put on biological and biogeochemical applications, especially in coastal regions.

Outlook: It can be envisioned that a combination of simulated (through numerical models) radiances with observed radiances will become a basic tool for merging and calibrating data from different satellite missions and to provide products of physical fields. While already substantially used in atmospheric applications, this approach is still in its infancy in satellite oceanography.

Bibliography

- Anderson, D., 2010. Early successes: El Nino, southern oscillation and seasonal forecasting. In Hall, J., Harrison, D. E., and Stammer, D. (eds.), *Proceedings of the "OceanObs'09: Sustained Ocean Observations and Information for Society" Conference*, 21–25 September 2009, Venice: ESA Publication WPP-306, Vol. 1.
- Balmaseda, M., Alves, O., Fujii, Y., Lee, T., Rienecker, M., Rosati, T., Stammer, D., Xue, Y., Freeland, H., and McPhaden, M., 2010. Role of the ocean observing system in an end-to-end seasonal forecasting system. In Hall, J., Harrison, D. E., and Stammer, D. (eds.), *Proceedings of the "OceanObs'09: Sustained Ocean Observations and Information for Society" Conference*, September 21–25, 2009, Venice: ESA Publication WPP-306, Vol. 1.
- Bell, M.J., Lefebvre, M., Le Traon, P.-Y., Smith, N., and Wilmer-Becker, K., 2009. *Oceanography* 22(3), 14–21, <http://dx.doi.org/10.5670/oceanog.2009.62>. - See more at: <http://www.tos.org/oceanography/archive/22-3.html#sthash.zRVBI6Zd.dpuf>.
- Bengtsson, L., Arkin, P., Berrisford, P., Bougeault, P., Folland, C. K., Gordon, C., Haines, K., Hodges, K. I., Jones, P., Kallberg, P., Rayner, N., Simmons, A. J., Stammer, D., Thorne, P. W., Uppala, S., and Vose, R. S., 2007. The need for a dynamical climate reanalysis. *Bulletin of the American Meteorological Society*, 88, 495–501.
- Giering, R., and Kaminsky, T., 1998. Recipes for adjoint code construction. *ACM Transactions on Mathematical Software*, 24, 437–474.
- Griffies, S. M., and Adcroft, A. J., 2008. Formulating the equations for ocean models. In Hecht, M., and Hasumi, H. (eds.), *Eddy Resolving Ocean Models*. Washington, DC: American Geophysical Union. Geophysical Monograph, Vol. 177, pp. 281–317.
- Griffies, S. M., Böning, C., Bryan, F. O., Chassignet, E. P., Gerdes, R., Hasumi, H., Hirst, A., Trguir, A., and Webb, D., 2001. Developments in ocean climate modelling. *Ocean Modelling*, 2, 123–192.
- Griffies, et al., 2010. In Hall, J., Harrison D. E., and Stammer, D. (eds.), *Proceedings of the OceanObs09: Sustained Ocean Observations and Information for Society. Conference* (Vol. 1), Venice, Italy, 21–25 September 2009. ESA Publication WPP-306.
- Haidvogel, D. B., and Beckmann, A., 1999. *Numerical Ocean Circulation Modeling*. London: Imperial College Press. Environmental and Management, Vol. 2.
- Köhl, A., and Stammer, D., 2008. Decadal sea level changes in the 50-year GECCO ocean synthesis. *Journal of Climate*, 21, 1876–1890.
- Lee, T., Stammer, D., Awaji, T., Balmaseda, M., Behringer, D., Carton, J., Ferry, N., Fischer, A., Fukumori, I., Giese, B., Haines, K., Harrison, E., Heimbach, P., Kamachi, M., Keppenne, C., Köhl, A., Masina, S., Menemenlis, D., Ponte, R., Remy, E., Rienecker, M., Rosati, A., Schroeter, J., Smith, D., Weaver, A., Wunsch, C., Xue, Y., 2010. Ocean state estimation for climate research. In Hall, J., Harrison, D. E., and Stammer, D. (eds.), *Proceedings of the "OceanObs'09: Sustained Ocean Observations and Information for Society" Conference*, September 21–25, 2009, Venice: ESA Publication WPP-306, Vol. 2 (Online). Available from World Wide Web: <http://www.oceanobs09.net/blog/?p=106>.
- Marotzke, J., Giering, R., Zhang, Q. K., Stammer, D., Hill, C. N., and Lee, T., 1999. Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity. *Journal of Geophysical Research*, 104, 29,529–29,548.
- Marshall, J., Hill, C., Perelman, L., and Adcroft, A., 1997. Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling. *Journal of Geophysical Research*, 102(C3), 5733–5752.
- Müller, P., 2006. *The Equations of Oceanic Motions*. Cambridge: Cambridge University Press. 291 pp.
- Olbers, D., Willebrand, J., and Eden, C., 2012. *Ocean Dynamics*. Berlin/Heidelberg: Springer, 704 pp.
- Reinicker, M., Awaji, T., Balmaseda, M., Behringer, D., Bell, M., Carton, J., Breivik, L. -A., Dombrowsky, E., Fairall, C., Freeland, H., Griffies, S. M., Haines, K., Ed Harrison, D., Heimbach, P., Kamachi, M., Kent, E., Lee, T., Le Traon, P. -Y., McPhaden, M., Martin, M. J., Oke, P., Palmer, M. D., Remy, E., Rosati, T., Schiller, A., Smith, D. M., Snowden, D., Stammer, D., Trenbert, K. E., and Xue, Y., 2010. Synthesis and assimilation systems: essential adjuncts to the Global Ocean Observing System. In Hall, J., Harrison, D. E., and Stammer, D. (eds.), *Proceedings of the "OceanObs'09: Sustained Ocean Observations and Information for Society" Conference*, September 21–25, 2009, Venice: ESA Publication WPP-306, Vol. 1.
- Stammer, D., Wunsch, C., Fukumori, I., and Marshall, J., 2002a. State estimation in modern oceanographic research. *Eos, Transactions, American Geophysical Union*, 83, 289, 294–295.
- Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C. N., and Marshall, J., 2002b. The global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model. *Journal of Geophysical Research*, 107(C9), 3118, doi:10.1029/2001JC000888.
- Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C. N., and Marshall, J., 2003. Volume, heat, and freshwater transports of the global ocean circulation 1993–2000, estimated from a general circulation model constrained by World Ocean Circulation Experiment (WOCE) data. *Journal of Geophysical Research*, 108, 3007, doi:10.1029/2001JC001115.
- Steward, R., 2005. *Introduction to Physical Oceanography* (Open Source Text Book). Available from World Wide Web: http://oceanworld.tamu.edu/resources/ocng_textbook/contents.html.
- Vallis, G., 2006. *Atmospheric and Oceanic Fluid Dynamics*. Cambridge: Cambridge University Press. 745 pp.
- von Storch, J.-S., Eden, C., Fast, I., Haak, H., Hernandez-Deckers, D., Maier-Reimer, E., Marotzke, J., and Stammer, D. 2012. An estimate of Lorenz energy cycle for the world ocean based on the 1/10° STORM simulation. *Journal of Physical Oceanography*, 42, 2185–2205, DOI:10.1175/JPO-D-12-079.1.
- Wunsch, C., 2006. *The Ocean Circulation Inverse Problem*. New York: Cambridge University Press. 442 pp.

Cross-references

Atmospheric General Circulation Models
 Climate Data Records
 Climate Monitoring and Prediction
 Global Climate Observing System
 Ocean Measurements and Applications, Ocean Color
 Ocean Surface Topography
 Ocean, Measurements and Applications
 Ocean-Atmosphere Water Flux and Evaporation
 Radar, Altimeters
 Sea Level Rise
 Sea Surface Salinity
 Sea Surface Temperature
 Sea Surface Wind/Stress Vector

OCEAN SURFACE TOPOGRAPHY

Lee-Lueng Fu

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

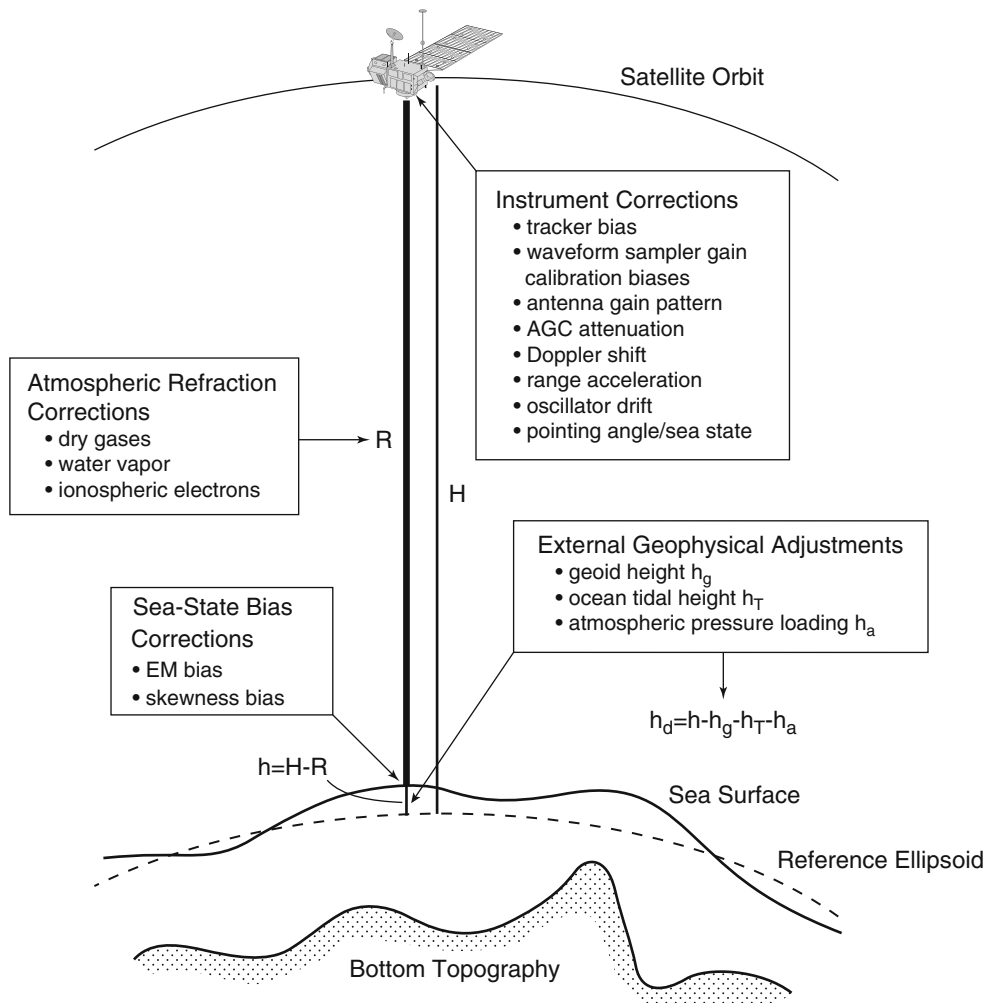
Definition

The deviation of the height of the ocean surface from the geoid is known as the ocean surface topography. The geoid is a surface on which the Earth's gravity field is uniform. The ocean surface topography is caused by ocean waves, tides, currents, and the loading of atmospheric pressure. The main application of ocean surface topography is for the determination of large-scale ocean circulation.

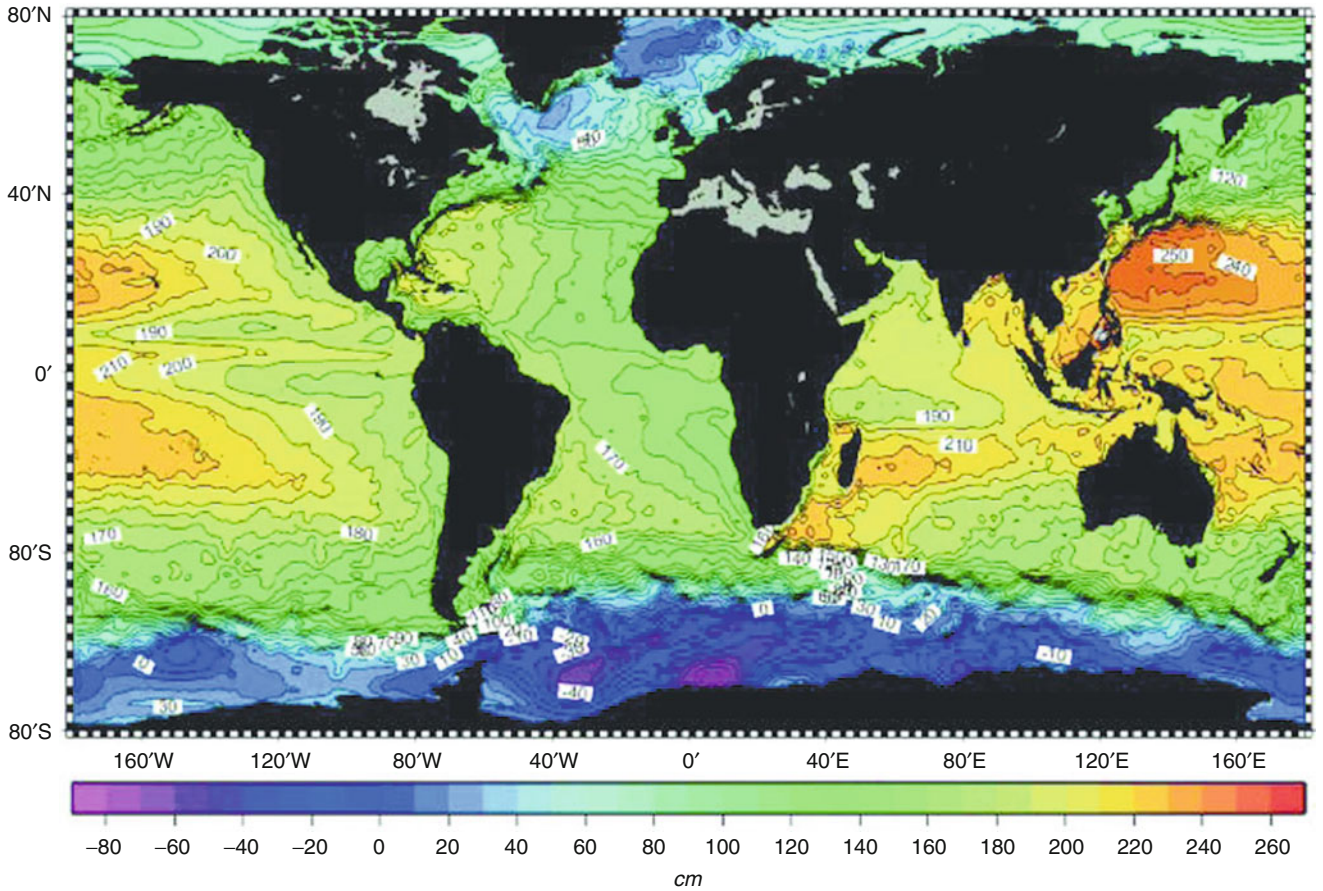
Introduction

The determination of global ocean circulation has been a challenging goal of physical oceanographers. The fluid

motion of the ocean has a wide range of spatial and temporal scales. We all have the experience of watching ocean surface and being awed by the ever-changing waves and their breaking into turbulence and white caps. In order to measure the velocity of an ocean current, we must average out these wave effects. At longer periods, there are ocean tides. If the timescales of our interests are longer than tides, we must also average out the effects of tides. Fortunately, there is a dynamic property of large-scale ocean currents called the geostrophic balance. This balance occurs when the Rossby number, defined as U/fL , is much smaller than unity. In the notation, U is the speed of the current, L is the spatial scale, and f is the Coriolis parameter defined as $f = 2\Omega \sin\phi$, where Ω is the Earth's rotation rate (7.292×10^{-5} rad/s) and ϕ is the latitude. For example, if $U = 50$ cm/s at 45° latitude, geostrophic balance occurs when $L \gg 5$ km. Then the geostrophic velocity of the current at ocean surface can be computed from the pressure at ocean surface as follows:



Ocean Surface Topography, Figure 1 Schematic illustration of the measurement system of satellite altimetry (From Chelton et al., 2001).



Ocean Surface Topography, Figure 2 OST derived from a combined analysis of satellite altimeter data, surface drifter data, and a geoid model (From Rio and Hernandez, 2004).

$$fv = \frac{1}{\rho} \frac{\partial p}{\partial x} \quad (1)$$

$$fu = -\frac{1}{\rho} \frac{\partial p}{\partial y} \quad (2)$$

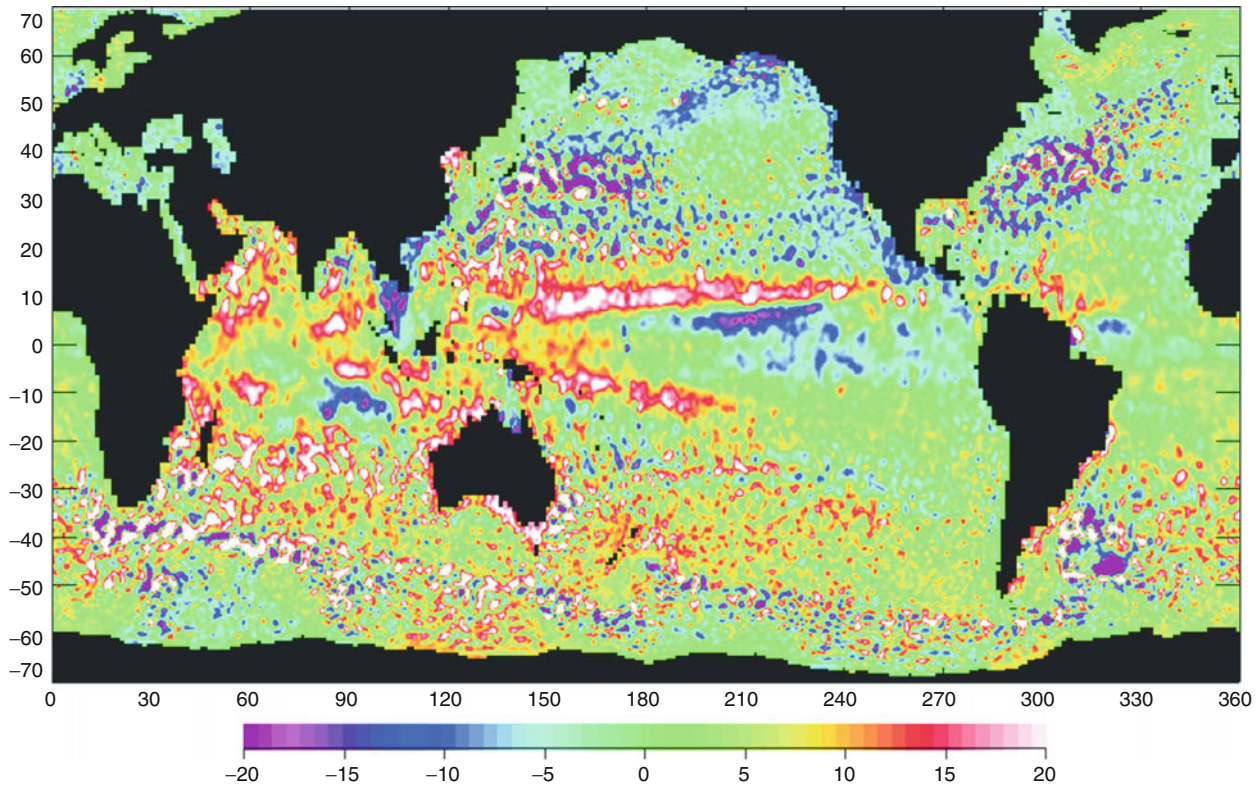
where p is the pressure, ρ is the density of sea water, and u and v are the zonal and meridional velocity components, respectively. The large-scale ocean currents are also in hydrostatic balance, namely, $p = g \eta$, where η is the elevation of the ocean surface topography (denoted by OST hereafter), defined as the deviation of the ocean surface from the geoid, and g is the Earth's gravity acceleration. The geoid is defined as the ocean surface on which the Earth's gravity field is uniform. In the absence of any motions, the ocean surface would be conformed to the geoid.

The role of OST for ocean currents is thus equivalent to the role of atmospheric surface pressure for winds. The speed and direction of large-scale wind blowing at

the Earth's surface can be computed from the gradient of the atmospheric surface pressure; similarly, the speed and direction of large-scale ocean surface currents can be computed from the gradient of OST as shown by Equations 1 and 2. OST has been traditionally estimated from the density field of the ocean as follows:

$$\eta_s = -\frac{1}{\rho_0} \left(\int_{-h}^0 \frac{\partial \rho}{\partial T} T' dz + \int_{-h}^0 \frac{\partial \rho}{\partial S} S' dz \right) \quad (3)$$

where T' and S' are the temperature and salinity deviations from their mean values, and ρ and ρ_0 are the density and its depth average, respectively, and h is the depth of the ocean. This quantity, η_s , has been traditionally referred to as the steric sea level, which is an approximation of OST. In most places, the steric sea level is predominantly determined by temperature, and therefore is a good indicator of the heat content of the water column of the ocean. OST is also influenced by other factors such as



Ocean Surface Topography, Figure 3 A snapshot (May 30, 2007) of the temporal change of OST from its time-mean value. The unit is cm.

wind, atmospheric pressure, and tides. However, our knowledge of OST has primarily been obtained through shipboard measurement of ocean temperature and salinity using Equation 3 until the recent development of satellite altimetry (Fu and Cazenave, 2001).

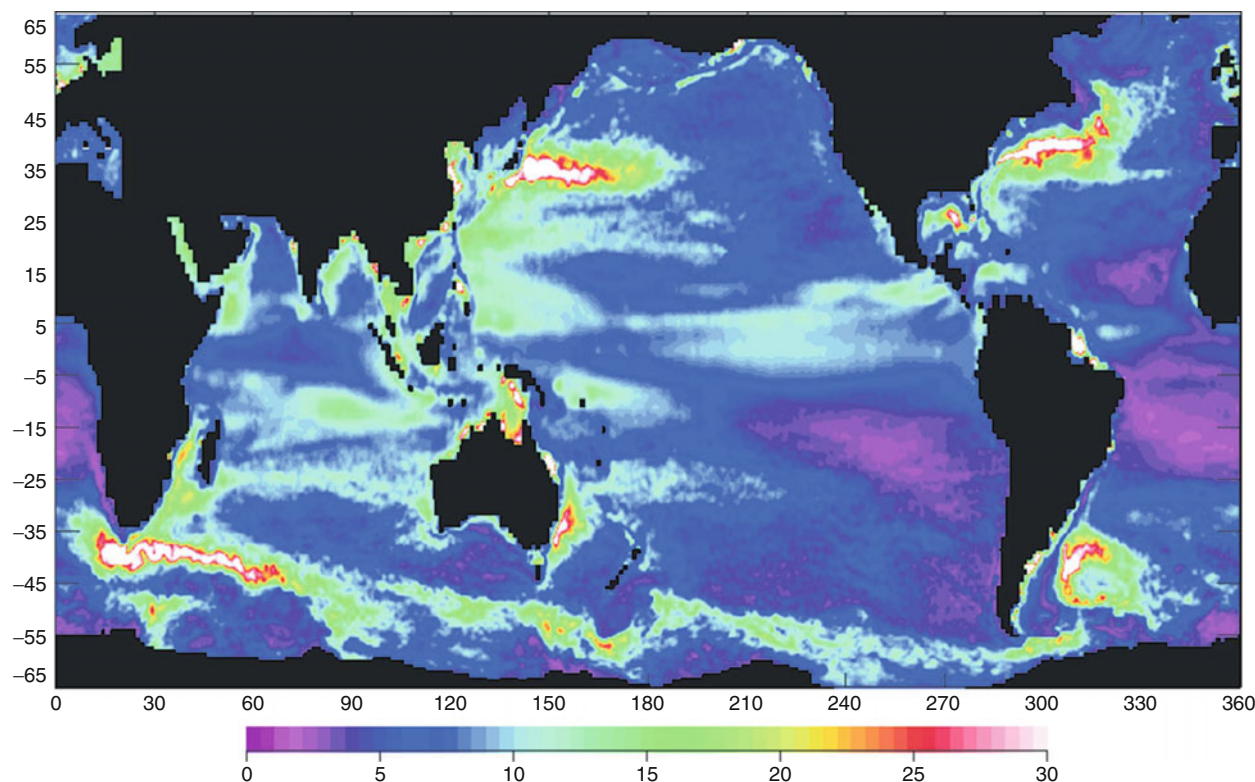
Satellite altimetry

The concept of using a radar altimeter for measuring the height of sea surface was developed in the 1960s. The measurement of the round-trip travel time of a radar pulse from a satellite to the surface of the ocean allows the determination of the distance between the satellite and the ocean surface. If the location of the satellite in orbit is also known in an Earth-fixed coordinate, then the height of the sea surface is determined in this coordinate and so is the OST if the knowledge of the geoid is available as well. The first satellite altimeter was carried by Skylab launched in the early 1970s. The first satellite altimeter with sufficient precision for detecting OST was onboard Seasat, which was launched in 1978 returning only about 100 days of data. However, the first image of the ocean reflecting the variability of OST caused by ocean currents provided a revolutionary impact on the

oceanography community. Oceanographers began to realize the potential of observing the global ocean circulation from OST determined from space.

The challenges of determining OST from satellite altimetry were the measurement accuracies. For example, an error of 1 cm in OST is capable of producing an error in the mass transport by ocean current of the magnitude of several million tons per second, which is a significant fraction of the transport of major ocean currents. The technique of satellite altimetry is conceptually straightforward, but there are numerous sources for measurement errors as illustrated in Figure 1. Corrections for these errors constitute the enterprise of precision altimetry beginning with the TOPEX/Poseidon Mission. The reader is referred to Chelton et al. (2001) for a treatise of the details. Only the major components of the measurement system are briefly described in this entry.

First, the free electrons in the Earth's ionosphere slow down the radar signals and have a net effect of more than 20 cm in error if not corrected. To correct for this error, radar altimeter must transmit in two channels, for example, the Ku band (13.6 GHz) and the C band (5.3 GHz) as the TOPEX/Poseidon altimeter. The signal delays are different for the measurements made at the two channels, which are used



Ocean Surface Topography, Figure 4 The standard deviation of the temporal change of OST in unit of centimeters.

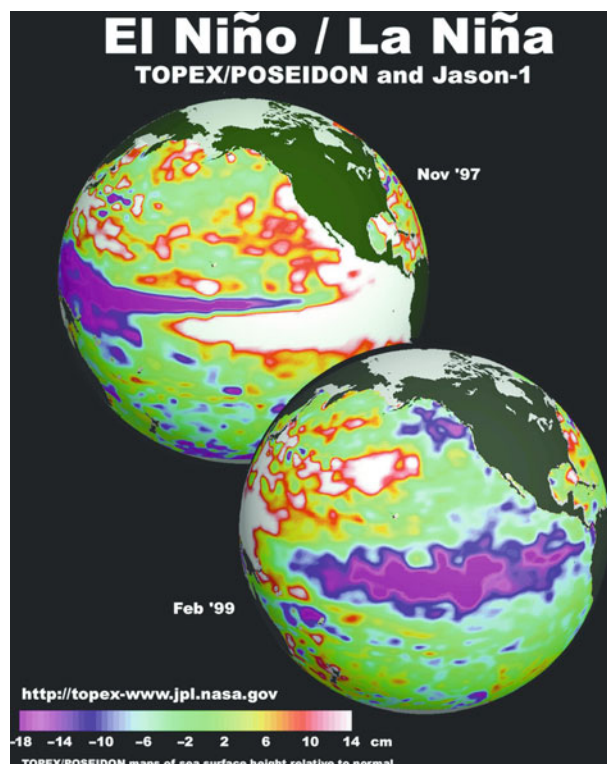
to determine the total electron content and the signal delay. Second, the water vapor in the Earth's troposphere also has the effect of delaying radar signals by more than 40 cm in the tropics. The approach of the TOPEX/Poseidon Mission was to carry a three-frequency microwave radiometer for making measurement of the water vapor content and the corresponding signal delay of the altimeter. Third, the precise location of the satellite in orbit must be determined. This requires onboard precision tracking systems and well-maintained geodetic ground systems. For example, the TOPEX/Poseidon Mission carries three independent orbit tracking systems: laser retroreflectors, GPS receivers, and a Doppler system called DORIS.

After Seasat, there were two satellite altimeters flown without the full suite of capabilities of error corrections: the US Geosat and the European Remote-Sensing Satellite (ERS). The first satellite altimeter mission that was dedicated to optimal measurement of OST was the joint American/French TOPEX/Poseidon Mission, launched in 1992. Aside from the complete measurement system, the 66° orbit inclination was chosen for the determination of ocean tides, and the 1336 km orbit altitude was chosen for precision orbit determination. The results from the mission have revolutionized the study of global ocean circulation from the OST determined from the mission and its follow-on, the Jason Mission, which was launched in 2001. Jason-2 was launched in 2008 to succeed Jason-1

to continue the precision altimetry data record. In the mean time, the other satellite altimeters from the ERS, ENVISAT, and Geosat Follow-On (GFO) missions have contributed to the measurement of sea surface height and OST by providing enhanced spatial and temporal resolutions. These missions have, however, benefited from the cross-calibration with the precision missions for correcting for large-scale errors such as orbit errors and tidal effects. It has been recognized by the altimetry community that the continuation of precision missions (e.g., Jason) in combination with other missions (e.g., ENVISAT) is essential for the measurement of OST for a variety of applications.

Ocean general circulation

The measurement of sea surface height from satellite altimetry is not sufficient for the determination of OST. The missing link is the geoid. Direct measurement of the Earth's gravity field for geoid determination is limited to a resolution of 300 km at present (Jayne, 2000). The OST derived from subtracting the geoid from the altimetry sea surface height data is thus lacking accurate information at smaller scales. In order to obtain OST with sufficient resolution to reveal the detailed features of the ocean general circulation such as the Gulf Stream and other ocean current systems, we have to supplement the satellite-derived result



Ocean Surface Topography, Figure 5 Maps of the temporal change of the OST of the Pacific ocean during the 1997 El Niño (upper globe) and the 1999 La Niña (lower globe).

with information derived from in situ measurements of ocean current velocity and density field. Displayed in Figure 2 is a map of time-averaged global OST estimated from combined data from satellite and in situ measurements (Rio and Hernandez, 2004). The spatial gradient of OST reveals the patterns of the ocean general circulation. The concentration of OST contours in major ocean current systems is clearly shown: the Gulf Stream in the western North Atlantic, the Kuroshio (or the Japan Current) in the western North Pacific, and the Antarctic Circumpolar Current around Antarctica. The data used for constructing the map were collected during the period 1993–1999, which is relatively short for representing the time-averaged ocean general circulation. Many of the details of OST and its gradient cannot be determined to be permanent features or manifestations of unaveraged temporal variability of ocean currents. Nevertheless, Figure 2 represents an estimate of the steady state of OST based on a state-of-the-art set of modern observations.

Mesoscale variability

Displayed in Figure 3 is an instantaneous map showing the temporal change of the OST from its time-averaged value shown in Figure 2. The map is constructed from a merged product from the altimeter data of the TOPEX/Poseidon and the ERS satellites

(Ducet et al., 2000). It is apparent that the spatial scales of the temporal variability in the tropics are much larger than those at the mid- and high latitudes, where the variability is dominated by the mesoscales with wavelengths shorter than 500 km. This mesoscale variability is the ocean's analog of the storms in the atmosphere. These mesoscale eddies are primarily created by the instability of ocean currents. They have maximum strength near the major current systems of the ocean, as illustrated by Figure 4 which is a map of the standard deviation of the OST calculated from the same merged data. The standard deviation of OST is a measure of the intensity of the ocean eddy field, which contains 90 % of the kinetic energy of ocean circulation and plays a significant role in transporting heat, salt, dissolved gases (e.g., CO_2), and nutrients around the global oceans.

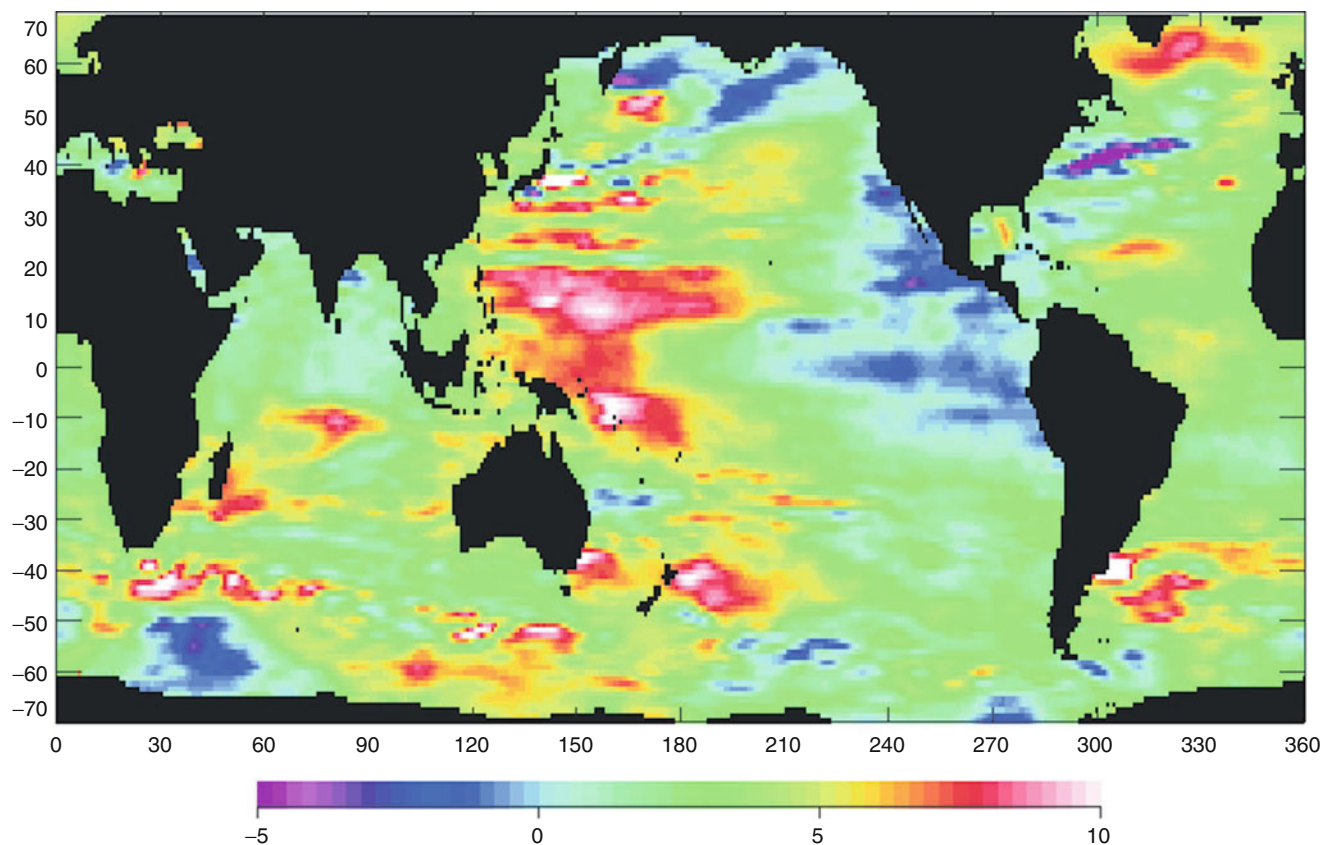
Basin-scale variability

The strongest variability at the scales of the ocean basins takes place in the tropics, as illustrated by the features in the Pacific Ocean and the Indian Ocean in Figure 3. In the Pacific Ocean, the tropical region is characterized by a seesaw pattern of exchange of heat between the western and eastern parts of the ocean resulting from an ocean–atmosphere coupled climatic system. When the heat content is anomalously high in the eastern and central tropical Pacific with OST higher than normal by up to 30 cm, the condition is called El Niño; when the heat content is anomalously low there with low OST, the condition is called La Niña (see Figure 5). The tropical Pacific is in these alternating states every 2–7 years with worldwide impact on weather and climate.

The basin-scale variability at mid- and high latitudes is quite different from that in the tropics. At high frequencies (with periods less than a season), such variability has a magnitude of generally less than 10 cm caused by the fluctuations of the volume of water mass in the entire water column driven by wind and atmospheric pressure at timescales of days to weeks. At low frequencies (with periods of a season and longer), the variability is associated with the seasonal, interannual, and decadal variability of the ocean. The forcing is a combination of thermal, hydrological, and wind origin. Displayed in Figure 6 is the geographic pattern of the linear trend of the temporal change in OST estimated from satellite altimetry data collected in the period from 1993 to 2008. The timescales of the variability are on the order of a decade, reflecting the climatic change of the ocean circulation and the global sea level. Note that the sea level in most regions of the global ocean is rising, causing a global mean sea level rise at a rate of 3.3 mm/year. Part of the rise is caused by the warming of the ocean that expands the volume of sea water, and the other part is caused by the melting of ice on land that adds water to the global oceans.

Practical applications

The information obtainable from OST on the circulation and the heat content of the ocean finds many practical



Ocean Surface Topography, Figure 6 The linear trend of temporal change of OST from 1993 to 2008. The unit is mm/year.

applications. For example, the fluctuations of ocean current velocity in the Gulf of Mexico have important effects on the safety and operational cost of the offshore oil drilling industry in the region. The data product of OST and current velocity has been made available to the offshore operators, resulting in cost savings of up to a half million US dollars in a single operation involving towing oil rigs along optimal routes. The upper ocean heat storage estimated from OST has proven useful for the prediction of the strength of hurricanes. This is because a hurricane draws its energy for growth from the heat stored below the sea surface. While the information of [sea surface temperature](#) reflects the heat only in a thin layer of the ocean surface, the heat stored below the ocean surface estimated from OST contains the fuel for hurricane to grow. Other applications include El Niño and La Niña predictions, fisheries management, marine mammal research, ship routing, and monitoring global [sea level rise](#).

Outlook

We have come a long way since the first global OST observed from Seasat. The observations from all the satellite altimeters have since revolutionized the fields of

oceanography and geodesy, as well as created many practical applications. However, there is a significant limitation in the conventional radar altimetry. The size of the pulse-limited radar footprint underlying the conventional radar altimetry increases from 2 to 10 km with the height of ocean waves. Even with thousands of pulses averaged over a second, the measurement noise over the large footprint has limited the spatial resolution of altimeter observation to wavelengths longer than 100 km. A major fraction of the ocean's kinetic energy is contained at wavelengths shorter than 100 km. Furthermore, the dissipation of ocean energy and the associated stirring and mixing of ocean properties like temperature and carbon dioxide content are controlled by physical processes at these short wavelengths. Many of the ocean processes in the coastal zones like the upwelling of nutrients, currents, fronts, and eddies that affect offshore operation, navigation, and waste disposal also have small scales that cannot be resolved by the conventional altimeter measurements.

A new technology of radar interferometry coupled with synthetic-aperture technique offers the opportunity of making high-resolution wide-swath measurement of OST (Fu and Rodriguez, 2004). This approach has been demonstrated by the Shuttle Radar Topography Mission

for mapping the land topography (Farr et al., 2007). By operating at a frequency above 30 GHz and a look angle of 4°, the concept of a wide-swath altimeter mission based on SAR interferometry has been proposed for mapping OST as well as the elevation of the water levels of rivers and lakes. A satellite mission with this capability will make breakthroughs in both ocean dynamics and land hydrology, which hold the key to understanding the evolving climate change as well as to monitoring and mitigating its consequences. Such a satellite mission called Surface Water and Ocean Topography (SWOT) has recently been established by a joint effort of the USA and France with contributions from Canada. SWOT is currently planned for launch in 2020.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Chelton, D. B., Ries, J., Haines, B., Fu, L.-L., and Callahan, P., 2001. Satellite altimetry. In Fu, L.-L., and Cazenave, A. (eds.), *Satellite Altimetry and Earth Sciences: A Handbook for Techniques and Applications*. San Diego: Academic, pp. 1–131. 423.
- Ducet, N., Le Traon, P. Y., and Reverdin, G., 2000. Global high resolution mapping of ocean circulation from the combination of TOPEX/POSEIDON and ERS-1/2. *Journal of Geophysical Research (Oceans)*, **105**(C8), 19477–19498.
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D., 2007. The shuttle radar topography mission. *Reviews of Geophysics*, **45**, RG2004, doi:10.1029/2005RG000183.
- Fu, L.-L., and Cazenave, A. (eds.), 2001. *Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications*. San Diego: Academic, p. 463. 463.
- Fu, L.-L., and Rodriguez, R., 2004. High-resolution measurement of OST by radar interferometry for oceanographic and geophysical applications. In Sparks, R. S. J., and Hawkesworth, C. J. (eds.), *State of the Planet: Frontiers and Challenges*. Washington, DC: American Geophysical Union. AGU Geophysical Monograph 150, IUGG, Vol. 19, pp. 209–224.
- Jayne, S. R., 2006. Circulation of the north Atlantic ocean from altimetry and the gravity recovery and climate experiment geoid. *Journal of Geophysical Research*, **111**, C03005, doi:10.1029/2005JC003128, 2006.
- Rio, M.-H., and Hernandez, F., 2004. A mean dynamic topography computed over the world ocean from altimetry, in situ measurements, and a geoid model. *Journal of Geophysical Research*, **109**, C12032, doi:10.1029/2003JC002226.

Cross-references

[Climate Data Records](#)
[Geodesy](#)
[Ocean Modeling and Data Assimilation](#)
[Ocean, Measurements and Applications](#)
[Radar, Altimeters](#)
[Sea Level Rise](#)

OCEAN SURFACE VELOCITY

Bertrand Chapron¹, Johnny Johannessen² and Fabrice Collard³

¹Satellite Oceanography Laboratory, IFREMER, Plouzané, France

²Nansen Environmental and Remote Sensing Center, Bergen, Norway

³CLS, Division Radar, Plouzané, France

Synonyms

Doppler; Surface velocity; Synthetic aperture radar

Definition

Doppler shift anomaly. Residual signal Doppler frequency from geometrically predicted Doppler frequency
Range Doppler velocity. Velocity conversion from Doppler shift anomaly estimate

Introduction

Regular repeat global measurements of absolute surface currents as well as ageostrophic processes including convergence and divergence along meandering fronts and eddies at scales smaller than 30 km are rare.

To derive global ocean surface current estimates from space, the basic tool is the radar altimeter. High-class altimeters (TOPEX, JASON, ERS, ENVISAT) are indeed measuring the oceanographic surface topography with good precision. Ocean surface current estimates are then retrieved assuming a linear balance between the gradient of the sea surface topography and surface geostrophic current. Satellite altimetry is a very mature satellite remote sensing technique with an extensive literature describing advances in knowledge and understanding of ocean circulation.

Satellite altimetry has in particular been suited to infer the statistical characterization of mesoscale variability. Thanks to its global, repeat, and long-term sampling, the estimation of sea surface height (SSH) wave number spectra has been a unique contribution of satellite altimetry to quantitative assessment of spectral slopes at scales ranging from about 50 to 100 km up to a spectral peak that generally emerges at twice the wavelength of the internal deformation radius (i.e., 300–400 km at mid-latitudes). Presently, assimilation of altimetry sea level anomaly (SLA) is routinely applied in several ocean modeling and forecasting systems. As such the availability and continuity of altimetric measurements are essential for operational oceanography. Yet, the finer-scale (less than 50 km) oceanic mesoscale and submesoscale energy is difficult to map with conventional radar altimeters because of the narrow illuminated swath, regardless of the orbital configuration.

Mesoscale and submesoscale processes are recognized for their importance and impact on biogeochemical processes and mixing, on air-sea interactions, and for the transfer of energy between scales. To adequately resolve

mesoscale eddies and ageostrophic processes and gradually fill this knowledge gap to advance the understanding of mesoscale and submesoscale processes, observations with finer space and time resolutions are consequently required. Unfortunately, clouds over the global ocean strongly inhibit regular use of very high 1 km resolution spaceborne optical sensors to delineate surface patterns and estimate surface advection. These observation deficiencies, in turn, are also manifested in the performance and quality of high-resolution ocean models. High-resolution observations are also obviously crucial for coastal applications for which the structure of surface currents is generally very complex, influenced by local bathymetry, tidal cycles, and wind and sea state conditions.

To complement spaceborne altimetric range measurements and to obtain direct surface velocity information from space, a different approach can be envisaged. Ideally, it is to use ocean surface waves as proxies to infer the mean ocean surface motion. Indeed, surface waves obey a dispersion relation, i.e., their phase velocities are wavelength dependent, linearly modified in the presence of a surface current. Such a principle is currently very successfully applied with in situ HF radar measurements. Anomalies between the theoretical phase speed of the probed ocean surface wavelength and the measured return signal frequency peak spectra are then interpreted in terms of surface currents.

From space, the interactions between electromagnetic and surface waves are less straightforward than for the near-grazing HF signals. More advanced technical developments and algorithms must then be invoked to exploit the measured Doppler information. Conventional along-track interferometry techniques have already been successfully implemented to measure the instantaneous sea surface scatter velocity by using the phase difference between two return signals from the same surface patch, separated by a very short time interval (Romeiser et al., 2005). In turn, this phase difference can be attributed to instantaneous sea surface scatter velocity. This is illustrated in Figure 1 showing an image acquired during the NASA Shuttle Radar Topography Mission (SRTM) over a complex and tidal-dominated coastal area off Brittany in France. The agreement with the range component of the 2D barotropic tidal current model is very good, both in magnitude and spatial pattern. The TerraSAR-X Synthetic Aperture Radar (SAR) mission launched in June 2007 and which has experimental along-track interferometry (ATI) capabilities provides a more systematic demonstration of this promising measurement technique for direct retrieval of surface current (e.g., Romeiser et al., 2010).

Direct, instantaneous frequency determination from the phase history analysis of single-antenna SAR returns is less conventional, but can also be used to retrieve a Doppler shift anomaly. As demonstrated by Chapron et al. (2005) and Johannessen et al. (2008), this Doppler anomaly is associated with an overall bulk velocity corresponding to the mean velocity of the radar-detected scatterers on the illuminated ocean surface, including a contribution from the ocean surface current. This

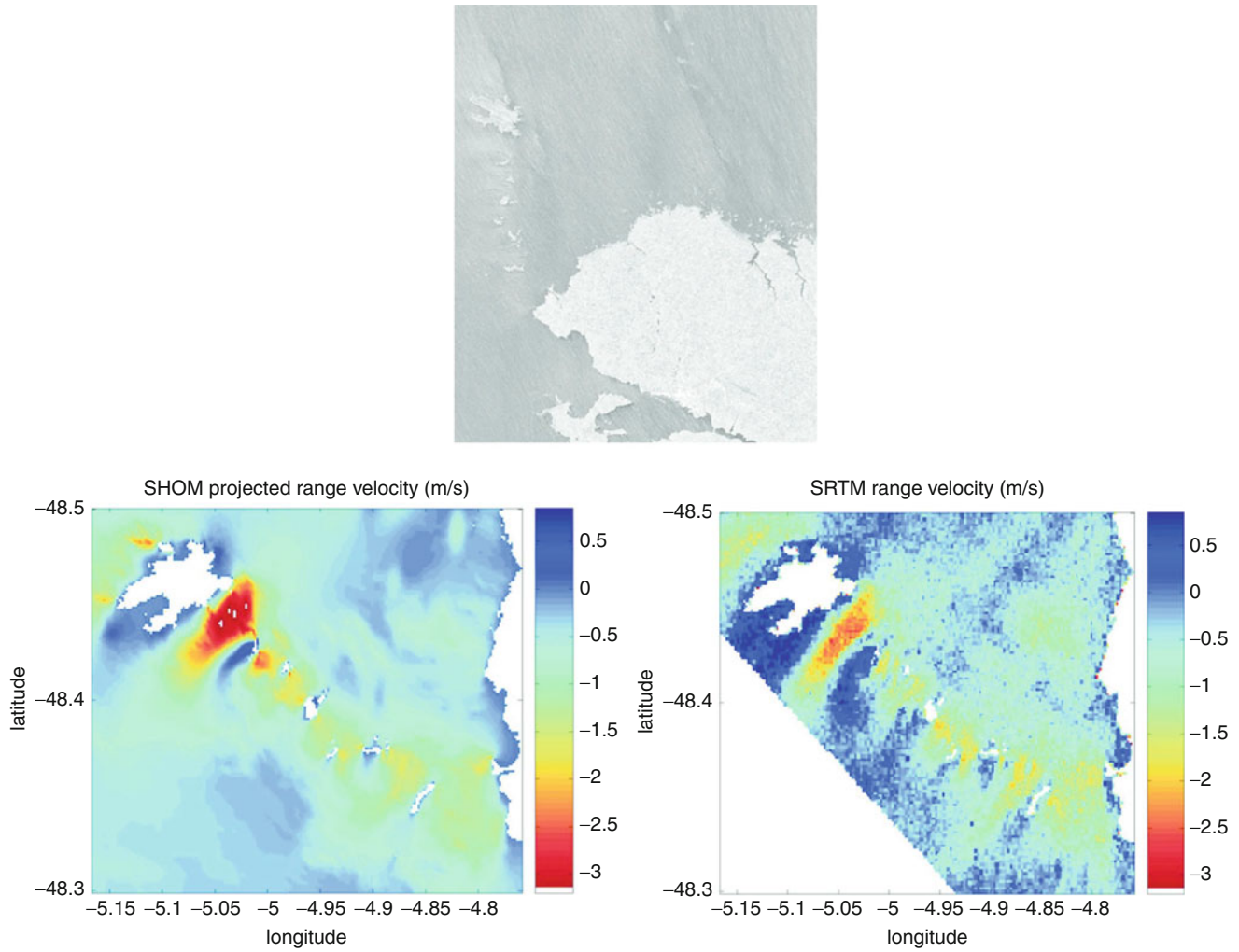
Doppler anomaly estimation is then dependent on the instrument characteristics (radar wavelength, incidence angle, polarization) and environmental conditions (mostly wind speed and direction). The partitioning of the signal to these contributing factors is very challenging, but the method has the potential to meet very high spatial resolution requisites. In all cases, the measurement of a Doppler frequency shift can complement the normalized radar cross section, so that both kinematic and dynamical properties of the ocean scene can be simultaneously derived.

Regular access to radar backscatter measurements and Doppler grid information from the ESA ENVISAT ASAR Wave Mode and Wide Swath Mode products has become feasible since mid-2007, thanks to ESA's upgrading of the ground segment. This has led to the gradual generation of a new and comprehensive data set for new explorations of spaceborne surface current measurements. In the next sections, the range Doppler shift signal retrievals are further examined and assessed in respect to (1) removal of the wind contribution, (2) the basin-scale equatorial surface velocity monitoring of the East Pacific, and (3) the strong and persistent surface velocity monitoring associated with the greater Agulhas Current. The subsequent section provides a summary of the key findings followed by a short outlook.

Range doppler velocity retrievals

Technically, single-antenna Doppler shift anomalies are obtained by subtracting the predicted from the measured return signal spectral peak frequencies, i.e., Doppler centroids. The method works best for homogeneous scenes, exhibiting small radar cross-sectional variations, and yields estimates with a resolution (azimuth, range) of about 10 km by 6 km for ENVISAT Advanced Synthetic Aperture Radar (ASAR) Wave Mode (WM) images and about 8 km by 4 km for ENVISAT ASAR Wide Swath Mode (WSM) images. For WSM products, prior to geophysical interpretations, corrections are applied to compensate along-track large cross-sectional variations and biases are further removed using land surface references. In order to use this method to derive quantitative information corresponding to the surface current, it is necessary to identify and remove the wind contribution to the measured Doppler anomaly.

Using a two-scale asymptotic decomposition of the surface roughness (or radar cross section), the radar imaging model (RIM) described by Kudryavtsev et al. (2005) provides estimations for the expected range Doppler scatter velocity by invoking a Doppler shift module following the approach outlined by Chapron et al. (2005). The Doppler shift results from the combined action of near-surface wind on shorter waves, longer wave motion, wave breaking, and surface current (Johannessen et al., 2008). Both kinematic (NRCS-normalized radar cross section) and dynamic properties of the moving ocean surface roughness can therefore be derived from the ENVISAT ASAR observations. Over an ocean imaged scene, the Doppler frequency f_D becomes a mean quantity:



Ocean Surface Velocity, Figure 1 (Top) NRCS measured from the SRTM mission. (Left) 2D tidal model current map. (Right) Coincident surface current map obtained from the SRTM interferometric technique (Courtesy P. Flament).

$$\frac{\pi f_D}{k_R} = - \frac{(u \sin \theta - w \cos \theta) \sigma_0(\theta + \Delta\theta)}{\sigma_0(\theta + \Delta\theta)} \quad (1)$$

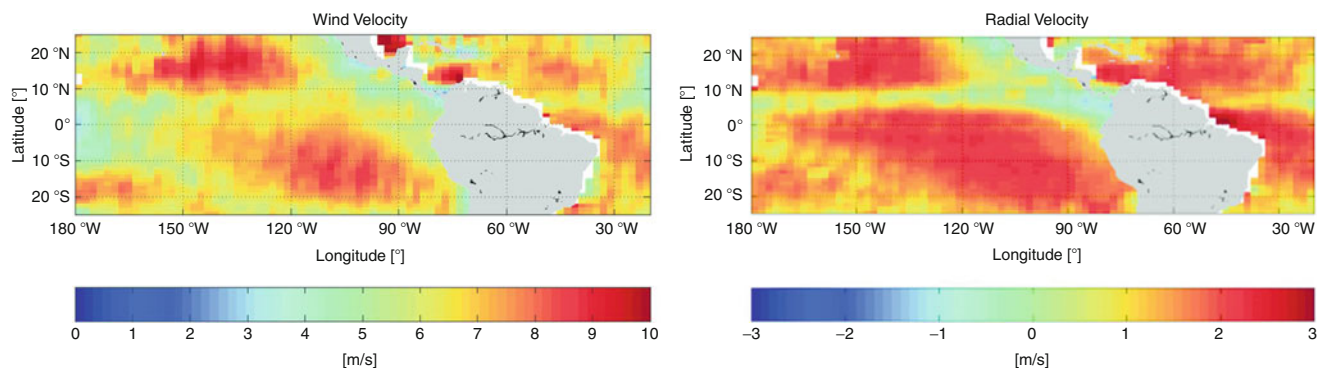
Here k_R is the radar wave number, u and w are the horizontal and vertical velocities of the scattering facets, and $\Delta\theta$ is the modification of the incidence angle θ due to the local tilt induced by the longer waves. The two-scale assumption helps to consider the NRCS variations caused by both the change of the local surface tilt ($\Delta\theta$) and the hydrodynamic modulation ($\tilde{\sigma}_0^h$) of the scattering facets, as $\tilde{\sigma}_0 = \Delta\theta \frac{\partial \sigma_0}{\partial \theta} + \tilde{\sigma}_0^h$, where $\Delta\theta = -(\cos \phi_R \cdot \zeta_x + \sin \phi_R \cdot \zeta_y)$, ϕ_R is the radar look direction, and ζ_x, ζ_y are the local components of the sea surface slope. Effects of surface tilt out of the incidence plane are ignored. To the second order in steepness, the RIM-derived radial velocity V_D of the target (assumed positive if directed away from the radar) then becomes

$$V_D = -\pi f_D / k_R \sin \theta = \bar{c}_f + u_s + c_f^{TH} \quad (2)$$

where $\bar{c}_f$ is the mean velocity of the scattering facets, u_s is the radial surface current velocity, and c_f^{TH} is the contribution due to tilting and hydrodynamic modulation of the facets. From the ability to partition the NRCS between regular scatter elements and breaking waves, the radial Doppler velocity in Equation 2 is further expressed as

$$V_D = u_s + \sum P_j^P(\bar{c}_j + c_j^{TH}) \quad (3)$$

where subscript j stands for Bragg waves (br), specular mirror points (sp), and breakers (wb). The specular point velocity always dominates V_D at low incidence angles (Mouche et al., 2008). With increasing incidence angles, this part of the Doppler velocity becomes negligible. At moderate incidence angles, the simulation of the total



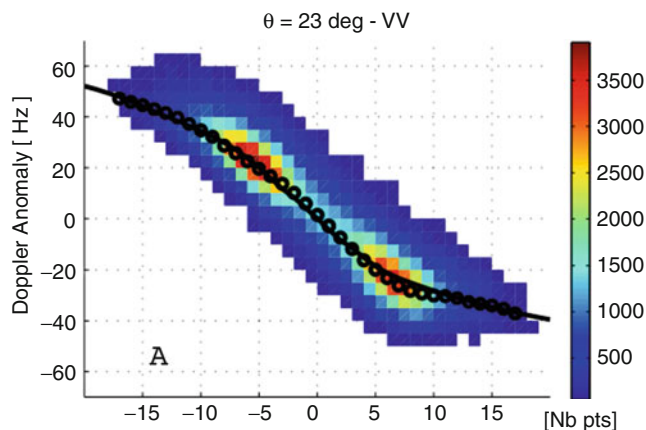
Ocean Surface Velocity, Figure 2 (Left) ECMWF monthly wind field projected along the radial direction of the ENVISAT ASAR antenna (ascending pass). (Right) Monthly range Doppler velocity estimated from the ASAR measurements. The average is performed for the month of November 2006 in boxes of 1° in latitude and 2° in longitude for the area centered to the east equatorial Pacific. Values are signed with respect to the orientation toward the SAR antenna so that radial velocity is positive (negative) when the wind blows with its radial component directed toward (away from) the antenna.

Doppler velocity predicts values that are about 35 % of the wind speed. This is significantly larger than expected from the phase speed of the Bragg waves and the wind-induced surface drift (about 3 % of wind speed). The two-scale decomposition with tilting and hydrodynamic effects explains this difference. Moreover, at moderate to large incidence angles, the breaking contribution cannot be neglected, and for HH, it eventually dominates V_D at very large angles. Further details of this approach are provided in Johannessen et al. (2008).

Estimating the wind contribution to the range doppler anomaly signal

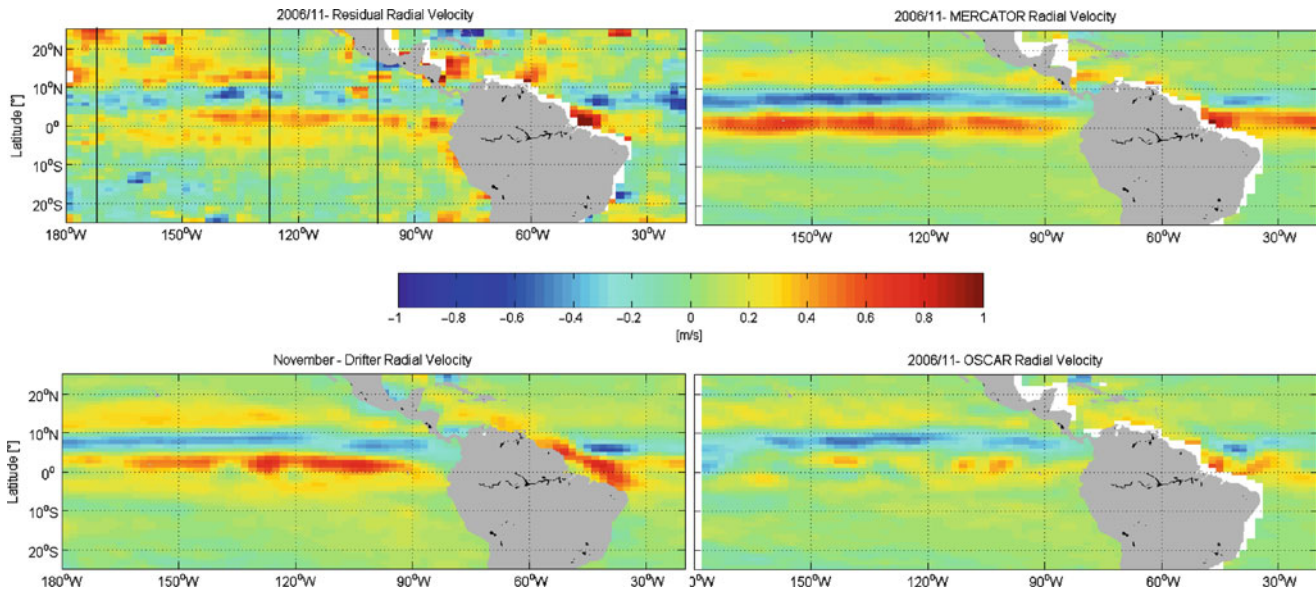
Under favorable and well-known wind conditions, inference of absolute surface velocities along the radar line-of-sight direction is highly feasible as suggested by Collard et al. (2008) and further applied in Johannessen et al. (2008). By taking full advantage of the large data set of the ENVISAT C-band ASAR collocated with ECMWF wind fields, Collard et al. (2008) developed a model relying on a neural network to predict the expected Doppler anomaly for a given wind speed. This empirical model thus allows systematic removal of the wind contribution to the measured Doppler anomaly, and the residual range Doppler velocity can be further assessed in view of surface current signatures.

At first order, the Doppler anomaly obtained using the Wave Mode products (at 23° incidence angle and VV polarization) is mostly wind dependent. This property of the Doppler anomaly is illustrated at a large scale in Figure 2 that presents the ECMWF monthly wind field projected along the radial direction of the ENVISAT ASAR antenna (Figure 2, left) and the corresponding monthly radial Doppler velocity estimated from the ASAR measurements (Figure 2, right). The correlation between radial wind given by ECMWF and Doppler radial velocity from ASAR is evident. Based on this collocated data set, the neural network model (called CDOP_23) is



Ocean Surface Velocity, Figure 3 Observed WM (color) and simulated (solid line) wind dependence of C-band Doppler shift VV polarization at 23° incidence angle. The color represents the spread in number of observation points. The open circles mark the mean fit to the observations. Upwind corresponds to positive radial velocity.

created, where the inputs are the wind and the relative direction of the wind with respect to the azimuth angle of the antenna. Furthermore, using the Doppler grid now available for each Wide Swath product, the CDOP_23 model has been extended to incidence angles from 17° to 42° . In so doing, a gradual decreasing trend in Doppler anomaly with increasing incidence angle is revealed. Such a result was anticipated by model simulations reported in Mouche et al. (2008) and Johannessen et al. (2008). In addition, a significant difference between up- and downwind directions is also encountered. As for the radar cross section, this difference increases with incidence, meaning that the relative distribution for small scatterers lying along longer waves is certainly not symmetric.



Ocean Surface Velocity, Figure 4 (Upper left) Residual radial velocity obtained from Doppler anomaly analysis. (Upper right) Radial velocity obtained using MERCATOR ocean circulation model, (lower left) monthly climatology available from the global drifter program, and (lower right) the ocean surface current derived from altimetry data and wind field analysis (OSCAR) Data and model fields are from November 2006.

As reported in Figure 3, the simulated Doppler frequency shifts display a functional relationship with wind speed in good agreement with the observations, in particular up to a wind speed of about 15 m/s. The simulated Doppler frequency shifts for VV polarization (and HH polarization, not shown) provide an acceptable shape versus wind speed. These results give strong confidence that the extended RIM simulations make an excellent analyzing tool to the ASAR NRCS and range Doppler velocity observations.

Ocean basin-scale range doppler velocity monitoring

Exploiting the larger number of ASAR Wave Mode observations and the development of the CDOP model, the detection capability of the Pacific equatorial current regime has been examined. The residual radial Doppler velocity obtained after the removal of the wind effect is presented in Figure 4 (upper left). It exhibits a band with significant easterly (negative) directed radial velocities around 7 °N latitude and two bands of westerly (positive directed) values on either side centered at 2 °N and 12.5 °N latitude. This latitudinal variation of the range-directed surface Doppler velocity is in general agreement with the expected positions of the equatorial current and the countercurrent.

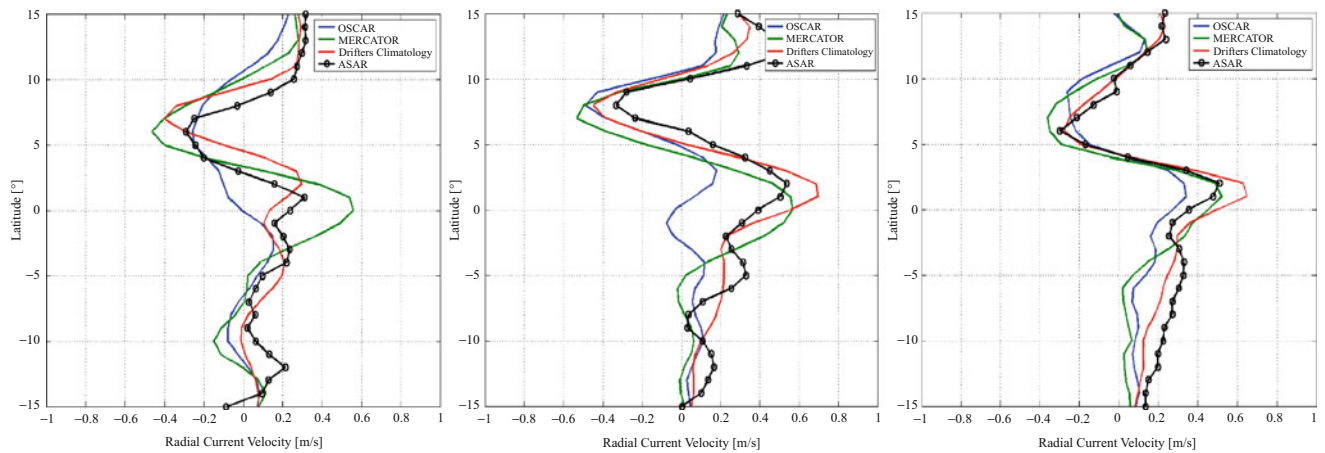
Assessment of this result is carried out by comparison to the sea surface current projected along the line of sight of the radar (Figure 4, upper right) given by the numerical global ocean circulation model MERCATOR (<http://www.mercator-ocean.fr/>), the drifter-derived climatology

of global near-surface currents (Figure 4, lower left) produced by NOAA in the framework of the global drifter program (Lumpkin and Pazos, 2007), and the ocean surface current derived from altimetry data and wind field analysis (Figure 4, lower right) from NOAA through the OSCAR project (Bonjean and Lagerloef, 2002). The overall agreement between the four radial velocity maps is noteworthy. The locations of the equatorial currents and countercurrents are predicted and measured at about the same latitudes, while the range of Doppler velocity values agrees with the sea surface current predictions and the independent measurements.

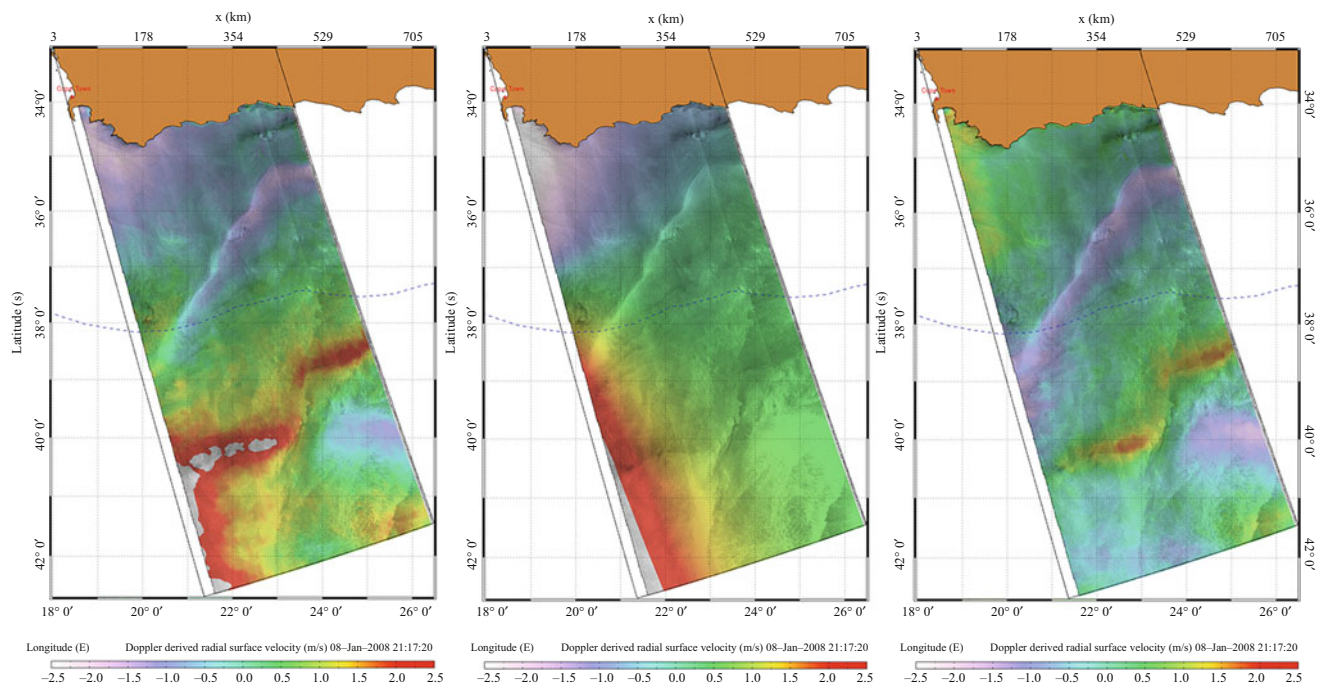
Further quantitative comparisons along three selected longitude bands are shown in Figure 5. Overall, the radial velocity from ASAR follows the variations of radial velocities given by the other sources. The equatorial countercurrent reaches up to 0.65 m/s and exceeds the equatorial current by 0.15 m/s. This is fairly consistent and with a difference that is slightly larger than the mean difference between the four individual sources which is less than 0.10 m/s. This suggests that both qualitatively and quantitatively, the sea surface current signature manifested in the Doppler anomaly from SAR Wave Mode data can be used, in combination with altimetry, surface drifters, and models, to advance the study of dynamics in the equatorial current regimes.

Regional Doppler range velocities: the Agulhas Current

The main flow directions of the Agulhas Current have an optimum configuration with respect to the line-of-sight



Ocean Surface Velocity, Figure 5 Comparisons of the radial velocity (m/s) given by MERCATOR (green), OSCAR (blue), drifters (red), and ASAR (black dotted) along the longitude bands marked in the upper left plot.

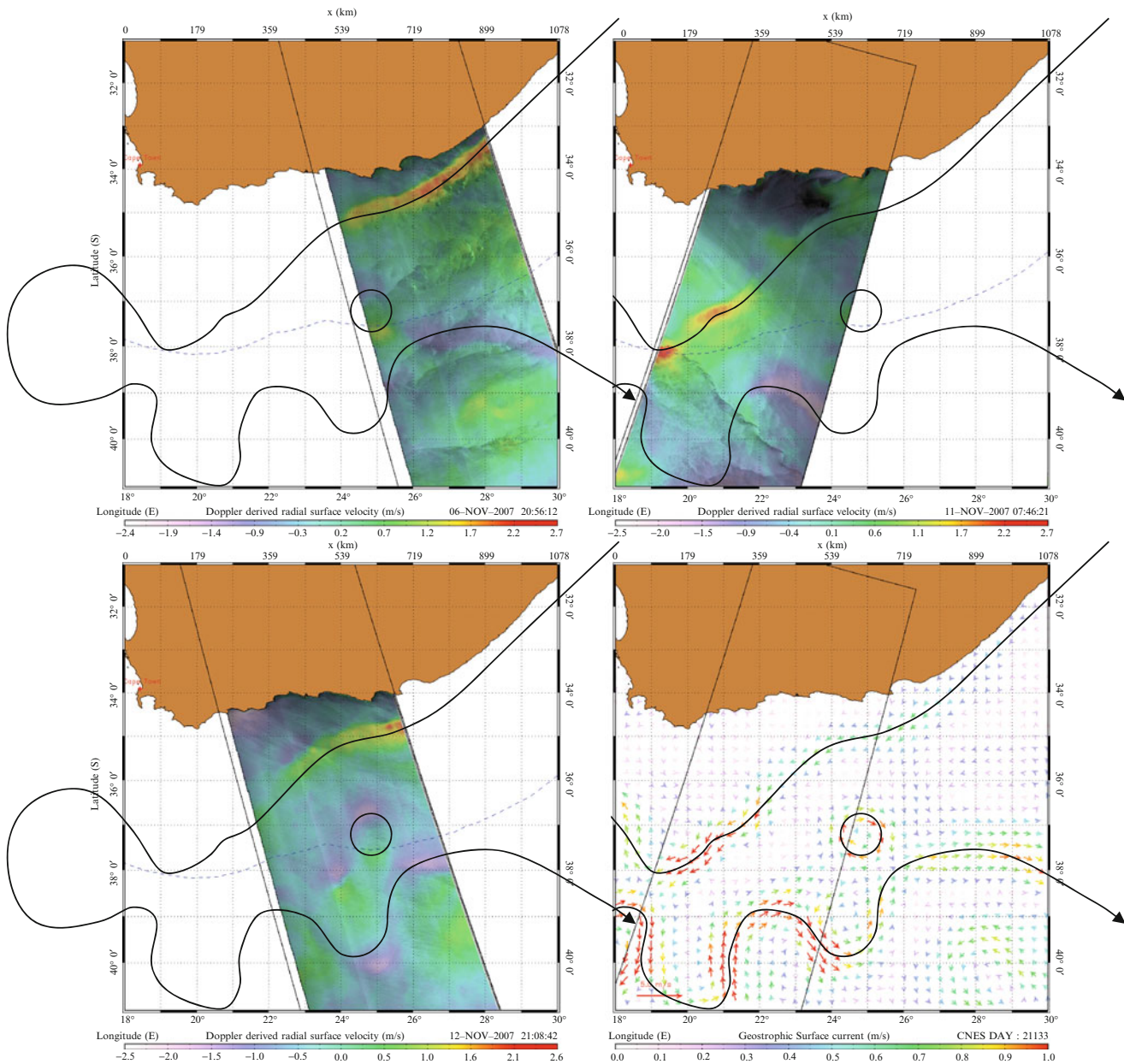


Ocean Surface Velocity, Figure 6 Illustration of the wind correction effect in the case of the Agulhas Current for January 8, 2008. (Left) Full range Doppler velocity. (Center) Wind contribution only. (Right) Range Doppler velocity after wind correction and removal.

velocity captured by the ASAR sensor for the ascending acquisitions as indicated in Figure 6 (left). Hence, after removal of the wind contribution to the signal, it becomes feasible to examine the residual velocity in connection with the strength and pathway of the greater Agulhas Current effect as demonstrated by Johannessen et al. (2008). After subtracting the estimated wind contribution (Figure 6, middle) using the empirical CDOP

relationship (Collard et al., 2008), the remaining range-directed Doppler velocity reaches 2 m/s in the Agulhas Current and nearly 1.5 m/s in the return current (Figure 6, right).

The large-scale surface current patterns are clearly visible, especially the core of the Agulhas Current and the return current. From this promising result, further evidence of the persistent range-directed surface



Ocean Surface Velocity, Figure 7 ASAR Wide Swath (420 km) Doppler velocity map from November 6, 2007 (*upper left*), November 11, 2007 (*upper right*), and November 12, 2007 (*lower left*). The *color bars* mark the radial Doppler velocities ranging from -2.4 m/s to $+2.7$ m/s. Positive speed is directed toward the SAR look direction. Weekly map of surface geostrophic current at 25 km resolution derived from radar altimetry from November 4 to 11, 2007, (*lower right*) with the rectangular box of the November 11 ASAR image inserted. The *color bar* marks the value of the surface geostrophic current ranging up to 1 m/s. The *black curve* marks the location of maximum geostrophic current derived from altimetry.

speed and pattern associated with the greater Agulhas Current can be assessed by the time series shown in [Figure 7](#) for acquisitions on November 6, 11, and 12, 2007. Note that these depict the signals after subtracting the wind contribution using CDOP. Forming part of the South-West Indian Ocean subgyre, the Agulhas Current flows poleward along the

southeastern coast of Southern Africa from 27°S with a speed of about 2 m/s, eventually retroflecting and meandering eastward back into the South Indian Ocean around 40°S at a reduced speed of around 1 m/s. This reversal of the mean flow translates into opposite sign radial surface velocities as indeed captured in the series of images.

The validation of such measurements remains a challenge as coincident reliable sea surface current measurements are very rare. However, using synoptic radar altimetry measurements, we can at least conduct a validation by comparison. The core position of the maximum surface geostrophic current derived from the mean 7 day (4–11 November) composite map of altimeter data superimposed on the Doppler velocity maps derived from ASAR is also shown in Figure 7. The mean location and flow direction of the southern part of the Agulhas Current and the beginning of the Agulhas Return Current agrees very well. However, the strength of the altimeter-derived surface geostrophic current appears suppressed. In particular, with respect to the Agulhas Current where the maximum geostrophic current is only about 0.6–0.7 m/s compared to the range-directed Doppler velocity that reaches more than 2 m/s, the latter speed is also reported from surface drifters trapped in the main current stream. Hence, this distinct expression of the Agulhas Current with a maximum speed around 2 m/s and a sharp shear of about 10^{-4} s^{-1} is predominantly reflecting the influence of the Agulhas Current on the Doppler velocity measurement. The same is valid for the Agulhas Return Current. Under favorable imaging geometry and reliable removal of the wind contributions, this demonstrates that it is possible to derive quantitative information of intense surface currents from the radial Doppler velocity. The ability to study surface current dynamics in combination with surface drifters and altimeter-derived surface geostrophic current should therefore advance.

Summary

Introducing direct observations of scatter velocity from space, SAR-derived Doppler anomalies, when corrected for wind effects, are found to regularly provide estimates of the surface velocity. With a persistent surface current flow direction versus the radar look direction, the greater Agulhas Current and the Pacific equatorial current become ideal for the examination of the Doppler shift measurements. Comparisons with estimates derived from surface drifters, geostrophic current derived from altimetry, and modeled surface current yield very promising results. The current regimes are depicted with greater detail than generally possible from altimetry, due to the geostrophic assumption, and from surface drifters due to inhomogeneous distribution. In particular, the SAR-derived range Doppler velocity seems to be able to retrieve the variation of the equatorial and counter-equatorial surface currents across the intertropical convergence zone. Similarly the full strength and persistence of the Agulhas Current with variable dominance of shear, convergence, and divergence zones are detected in contrast to altimetry that tends to suppress the strength of the highly topographically steered Agulhas Current (Gründlingh, 1983).

In summary, the results are considered promising for strengthening the use of SAR in quantitative studies of the ocean currents. Combined with surface drifters and

altimeter-derived surface geostrophic currents, monitoring of the dynamics of intense current regimes may certainly be advanced. However, promising and exciting as these results are, the SAR-based Doppler velocity retrievals are in need of careful validation. A dedicated field campaign is therefore highly recommended. Execution of a campaign would also be very timely in view of the Sentinel-1 SAR mission, expected to be launched in 2014, to ensure continuation of C-band SAR data from ESA's ERS-2 and ENVISAT satellites. Important applications driving Sentinel-1 include marine vessel detection, oil spill mapping, and sea ice monitoring. With adequate validation and retrieval accuracy of the new method and products reported here, Sentinel-1 could be expected to provide additional consistent information on winds, waves, and currents.

Acknowledgments

This work was supported by the Norwegian Space Centre and the European Space Agency (ESA) through the Prodex Arrangement with contract no. 90266 and by the Research Council of Norway, project number 177441/V30. Support by ESA through the study contract no. 18709/05/I-LG, Service Hydrographique et Océanographique de la Marine (SHOM) through contract 05.87.028.00.470.29.25, and INTAS Association through the INTAS-06-1000025-9264 project are also acknowledged. We are also grateful for the altimeter data obtained from www.aviso.oceanobs.com, AVISO-Centre National d'Etudes des Spatiales (CNES).

Bibliography

- Bonjean, F., and Lagerloef, G. S. E., 2002. Diagnostic model and analysis of the surface currents in the tropical Pacific ocean. *Journal of Physical Oceanography*, **32**(10), 2938–2954. <http://www.oscar.noaa.gov/>.
- Chapron, B., Collard, F., and Ardhum, F., 2005. Direct measurements of ocean surface velocity from space: interpretation and validation. *Journal of Geophysical Research*, **110**, C07008.
- Collard, F., Mouche, A., Chapron, B., Danilo C., and Johannessen, J. A., 2008. Routine high resolution observation of selected major surface currents from space. In *Proceedings ESA Seasat 2008 workshop, ESA/ESRIN*, January 21–25, 2008. ESA Publication Division, Frascati.
- Gründlingh, M. L., 1983. On the course of the Agulhas current. *South African Geographical Journal*, **65**(1), 49–57.
- Johannessen, J. A., Chapron, B., Collard, F., Kudryavtsev, K., Mouche, A., Akimov, D., and Dagestad, K.-F., 2008. Direct ocean surface velocity measurements from space: improved quantitative interpretation of Envisat ASAR observations. *Geophysical Research Letters*, **35**, L22608.
- Kudryavtsev, V., Akimov, D., Johannessen, J. A., and Chapron, B., 2005. On radar imaging of current features. Part 1: model and comparison with observations. *Journal of Geophysical Research*, **110**, C07017.
- Lumpkin, R., and Pazos, M., 2007. Measuring surface currents with SVP drifters: the instrument, its data, and some recent results. In Griffa, A., Kirwan, A. D., Mariano, A., Ozgokmen, T., and Rossby, T. (eds.) *Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics*, Chapter 2. Cambridge University Press. http://www.aoml.noaa.gov/phod/dac/drifter_climatology.html.

- Mouche, A. A., Chapron, B., Reul, N., and Collard, F., 2008. Predicted Doppler shifts induced by ocean surface wave displacements using asymptotic electromagnetic wave scattering theories. *Waves in Random and Complex Media*, **18**(1), 185–196, doi:10.1080/17455030701564644.
- Romeiser, R., Breit, H., Eineder, M., Runge, H., Flament, P., de Jong, K., and Vogelzang, J., 2005. Current measurements by SAR along-track interferometry from a Space Shuttle. *IEEE Transactions on Geoscience and Remote Sensing*, **43**(10), S2315–S2324.
- Romeiser, R., Suchandt, S., Runge, H., Steinbrecher, U., and Grünler, S., 2010. First analysis of TerraSAR-X along-track InSAR-derived current fields. *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 820–829.

OCEAN, MEASUREMENTS AND APPLICATIONS

Ian Robinson

Ocean and Earth Science, University of Southampton,
at National Oceanography Centre, Southampton, UK

Synonyms

Oceanography from space; Satellite oceanography

Definition

Oceanography. The scientific study of the ocean and its contents, including the biology, chemistry dynamics, and physics of the sea and the geological and geophysical processes affecting the sea floor.

Impact of remote sensing on oceanography

Introduction

It is well over four decades since oceanographers first started to consider what they might learn about the ocean by observing it from above, using sensors on high-altitude platforms. It is remarkable how creative and farseeing their early proposals were (Ewing, 1965). Long before low-power electronics, computer-based digital signal processing, and high-capacity data storage devices were technologies familiar to environmental scientists, the first pioneers envisaged temperature- and phytoplankton-mapping radiometers, altimeters for determining ocean currents, and radars to measure wind, waves, and other phenomena that create sea surface roughness patterns. It took another 15 years, during which men first landed on the moon, before it was possible to prove the feasibility of these basic concepts (Gower, 1981; Saltzman, 1985). This success followed the launch in 1978 of three satellites (Seasat, TIROS-N, and Nimbus-7) which convincingly demonstrated that sensors on satellites could provide ocean scientists with measurements of ocean properties that revealed new insights and changed the way they perceived the ocean.

By 1985 the subject of satellite oceanography was becoming established, with the publication of three textbooks (Maul, 1985; Robinson, 1985; Stewart, 1985)

outlining the principles and methods for measuring ocean properties from satellite data. These have not changed fundamentally, although more recent general textbooks (Martin, 2004; Robinson, 2004) and technique-specific monographs (e.g., Cracknell, 1997; Halpern, 2000; Fu and Cazenave, 2001; Gade, 2005) show improved accuracy and reliability of sensors and twenty-first century satellite data being used for a remarkable range of applications to diverse branches of oceanography (Robinson, 2010).

This entry first considers the factors in satellite oceanography that provide a unique view of the ocean. It outlines the main sensor classes, the remote sensing techniques associated with them, and the different ocean parameters that each can measure. It summarizes the types of ocean phenomena which are observed, emphasizing where satellite data have opened up new fields of oceanographic research. It also notes the more recent development of operational applications of satellite data for warning of hazardous sea conditions, supplying observations that constrain forecasting models, and monitoring the climate of the ocean. It identifies some challenges for future development and reflects on the overall impact that remote sensing has made on our knowledge of the ocean. Although airborne or land-based remote sensing methods are sometimes used to view the sea, their impact has been much less than satellites and they are not discussed here.

The unique ocean perspective offered by remote sensing

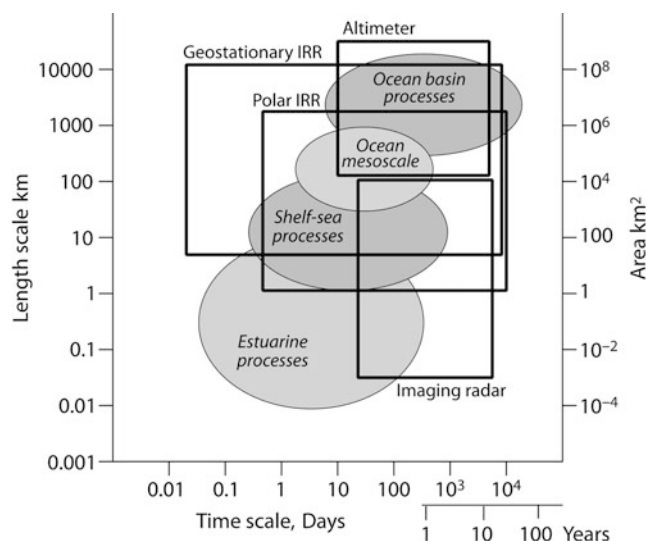
Earth orbiting satellites offer a number of unique advantages as platforms for ocean-viewing sensors. First is their capacity to acquire spatially detailed views of the ocean surface. Their typical height of 700–900 km in polar orbit, or 36,000 km in geostationary orbit over the Equator, gives them a wide synoptic view over the surface of the planet. Typical scanning sensors on polar-orbiting satellites resolve spatial detail down to about 1 km, although the resolution of different types of ocean-viewing instruments ranges from 50 km down to 25 m. There is a practical limit of between 1,000 and 4,000 for the ratio between the linear extent of a synoptic image and its spatial resolution. Thus a scanning radiometer with a nadir resolution of 1 km typically has a swath width of 1,000–2,000 km, but beyond the swath extremities, the view is too oblique to be useful. Given the dynamic nature of the ocean and the way in which its surface properties are continually changing, conventional in situ ocean observing instruments lack the sampling capability to capture a nearly instantaneous spatially detailed “snapshot” of the two-dimensional distribution of ocean surface properties. The capacity of satellite data to do so has given the oceanographer a previously inaccessible perspective, which accounts for the pervasive impact that remote sensing has had on ocean science in recent years.

A second unique advantage of remote sensing is the global coverage of polar-orbiting satellites. This allows

data to be sampled in distant and dangerous waters such as the Southern Ocean just as easily as places more readily sampled using conventional ocean observing techniques. Furthermore, sensors in polar orbit provide consistent measurements across the world ocean. Within the limitations of particular methodologies, oceanographers can make confident comparisons between ocean measurements in different oceans or at different latitude made by a single sensor, without the need for complex intercalibration of the many different in situ instruments that would be needed to replicate the global reach of satellites. This capacity has helped ocean science to evolve into the study of basin-scale and global-scale phenomena in the last two decades.

A third significant ocean sampling capability of satellite observing systems is the repeated acquisition of data over many years. While the revisit interval for polar-orbiting sensors places a lower limit on the sampling period of 12 h, or several days if sensors have a fine spatial resolution (Robinson, 2004, pp. 72–75), they continue to sample regularly for several years over the lifetime of a sensor. Where data are acquired as part of an operational service, replacement satellites have maintained the collection of global-coverage time series for well over two decades in some cases. This has given oceanographers a new opportunity to study long-lived phenomena and especially, by extracting mean or seasonally varying climatologies from a longtime series, to immediately detect any anomalous behavior of the ocean relative to the climatology.

The space-time sampling capabilities of some of the main classes of ocean observing sensors are shown as rectangular boxes in Figure 1. The bottom left corner of each box represents on the y-axis the spatial resolution of the sensor or the pixel size (whichever is the larger) and on the x-axis the sampling interval assuming no data losses due to cloud or other factors. The top boundary represents the spatial size of the synoptic view from a single overpass and the right boundary is the length of the time series acquired. It is evident that there is a trade-off between spatial and time resolution, and the ratio between spatial resolution and near-instantaneous coverage is shown by the height of each box, given the logarithmic axes of the figure. The area covered by each box encompasses the range of length and timescales which the sensor is capable of detecting. In all cases, this far exceeds the capacity of an in situ array of moorings, drifting buoys, or vessels in transit. By plotting the sampling capabilities of sensors in this way, comparison can be made with the space and timescales of ocean phenomena, which are shown as shaded ellipses in Figure 1. In general, the short-length-scale phenomena have short timescales and larger phenomena have longer timescales. If ocean processes are to be effectively observed, it is important to match the space-time sampling capacity of instruments to the space-time variability and extent of the ocean phenomenon. Figure 1 shows that satellite data are very well matched to the observation of ocean processes at the basin scale down to the scale of



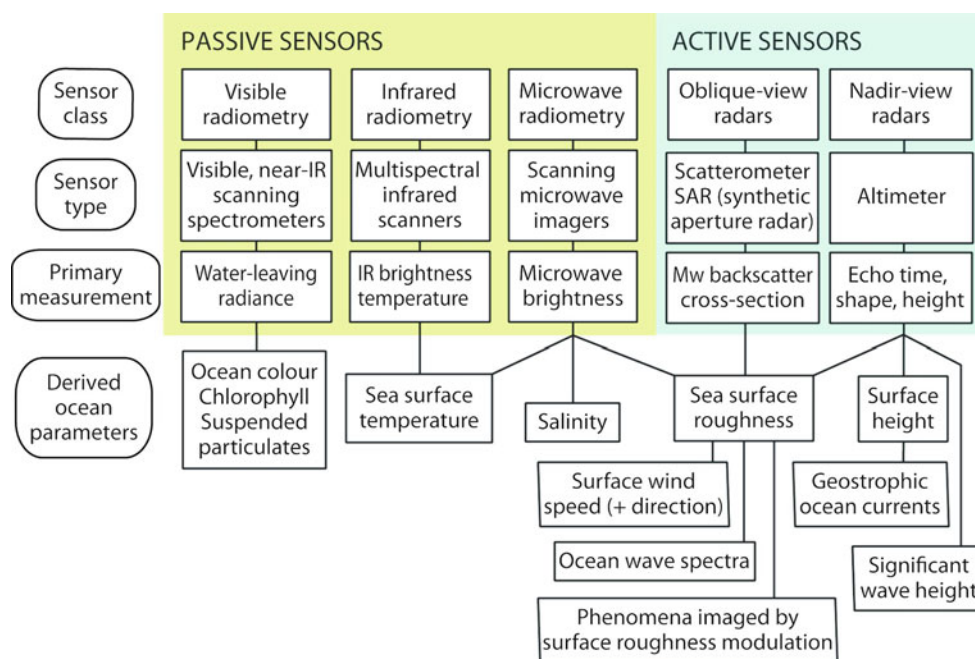
Ocean, Measurements and Applications, Figure 1 Schematic showing in rectangular boxes the space-time sampling capabilities of four different sensor types representative of ocean remote sensing instruments. Note that “polar” and “geostationary” refer to the orbit of the satellite platform and IRR stands for “infrared radiometer.”

shelf seas. Crucially there is a very good match to the ocean mesoscale, which is the scale of geostrophic turbulence in the ocean, typified by ocean eddies. Failure by in situ instruments to sample this scale properly not only makes it difficult to detect important ocean phenomena without ambiguity but can also lead to aliasing when trying to monitor the large-scale ocean circulation. Figure 1 also makes it clear that satellites are not able to sample adequately the variability scales of estuarine and coastal sea phenomena. Although fine resolution sensors such as imaging radars or land mapping instruments can acquire occasional snapshots of phenomena down to tens of meters, it is normally several days before the next view can be obtained by which time the phenomenon has probably disappeared and so its evolution cannot be tracked.

Fundamentally it is the near-perfect match of satellite sampling capacity to the ocean mesoscale, as well as the global and basin scale coverage, that has given remote sensing such an important role in modern oceanography, irrespective of which ocean variable is being measured. This impact has been made in spite of the other main limitation of remote sensing methods, that they cannot penetrate far, if at all, below the sea surface.

Methods and measurements of ocean remote sensing

In order to grasp the scope of ocean remote sensing, it is important to distinguish between the different classes and types of remote sensing instruments, the primary physical properties that they detect, the oceanographically relevant variables that can be retrieved from the primary



Ocean, Measurements and Applications, Figure 2 Schematic showing the different classes and types of satellite sensors used for viewing the ocean, the primary measurement that they make, and some of the ocean properties that are derived.

property, and the ocean phenomena that are revealed by the measured variables. Figure 2 makes this distinction in summarizing the different classes of sensor used for satellite oceanography, what they individually detect, and the ocean measurements that can be derived. The ocean phenomena that can be observed will be mentioned later.

Sensors used to observe the ocean

Ocean remote sensing methods (Robinson, 2004, pp. 11–26) use three different sources of electromagnetic radiation (EMR) to communicate information about the sea from its surface to a sensor in Earth orbit above the atmosphere. These are the following: (a) solar radiation reflected at or below the sea surface in the visible and near-infrared wavelengths of the EMR spectrum, (b) thermal radiation emitted by the sea surface in both the thermal infrared and the microwave, and (c) pulses of microwave radiation generated on the satellite, emitted toward the Earth, and then reflected by the sea surface. Passive sensors use (a) and (b), whereas active sensors are needed to use (c). As shown in Figure 2, the sensors and main methods of ocean remote sensing can be classified by the part of the spectrum being used and whether the sensor is active or passive. Passive sensors operate in the three principal spectral windows through which EMR passes through the atmosphere, that is, the visible and near-infrared waveband, the thermal infrared, and the microwave. Active sensors on satellites all operate in the microwave and so are types of radars. So far no lidars, active sensors operating in the visible waveband, have

been used for ocean remote sensing from satellites, although they are used from aircraft. Detailed information about the main types of sensor is found under their separate entries. Their capabilities are outlined here to show the wide scope of ocean remote sensing methods.

Visible waveband radiometers measure radiance at the top of the atmosphere (TOA), sampled in discrete spectral bands across the visible and near infrared in order to describe its spectral composition or “color”; hence, these instruments are typically referred to as “ocean color sensors.” The primary variable delivered by an ocean color radiometer system is the water-leaving spectral radiance, although this differs significantly from what is detected by the sensor above the atmosphere. Atmospheric correction (Gordon, 1997) is a major component of the primary processing required to retrieve the ocean color (i.e., the water-leaving radiance). The near-infrared channels are essential for this atmospheric correction procedure. There is almost no component of near-IR in the water-leaving radiance and so any detected in the TOA signal provides a measure of the amount of atmospheric path radiance that is present.

Ocean color in itself is not of particular interest for most marine scientists, but it is influenced by a number of different components which determine the water’s inherent optical properties. Most important of these is the concentration of the pigment chlorophyll-a associated with the presence of phytoplankton and with primary production. Chlorophyll-a concentration is the main ocean variable derived from ocean color sensors, although other factors also influence the color such as dissolved organic material

and suspended particulates. In the open ocean, these other factors are correlated with chlorophyll, which strongly absorbs blue light around 440 nm, giving the water a characteristic green color. This forms the basis of geophysical algorithms to retrieve the chlorophyll-*a* concentration from the green/blue spectral ratio in the water-leaving radiance (O'Reilly et al., 1998). However, where the other optically important water constituents have a source independent of the local phytoplankton population, which often happens in shallow and coastal seas, it is much harder to retrieve chlorophyll-*a*. This remains a major challenge for ocean color remote sensing. Despite this, image datasets of chlorophyll are produced daily from ocean color sensors at a resolution of 1–2 km, allowing detailed mapping of phytoplankton blooms. A record of weekly global composites over 10 years is now available to serve the needs of the international ocean biology and biogeochemistry community (McClain et al., 2006).

The next class of sensor in Figure 2, *passive infrared radiometers*, measures the thermal emission from the sea surface skin (approximately the top 10 μm), after it has passed through the atmosphere and been slightly absorbed by water vapor and other “greenhouse” gases. The primary measurement is the TOA brightness temperature (the temperature of a black body emitting the same radiance as that measured by the radiometer), and the main derived ocean variable is the [sea surface temperature \(SST\)](#). To retrieve this requires atmospheric correction, which relies on sampling multiple spectral bands across the thermal IR, since the water vapor absorption is wavelength dependent (McMillin and Crosby, 1984). Stratospheric aerosols also scatter and absorb the radiation, but these are less reliably detected, unless a dual view strategy is adopted, which observes the same location twice, at different incidence angles and hence through different atmospheric path lengths (Brown et al., 1997). There are several different platforms in near-polar orbits carrying IR radiometers and delivering SST maps that cover the globe with a resolution of 1–2 km at least daily. Sensors on geostationary platforms deliver SST maps, restricted to the horizon from the platform, at 4–5 km resolution every hour. Satellite SST data have a wide variety of ocean applications since many phenomena influence the surface temperature, creating a “signature” for the phenomenon in the spatial distribution of SST. However, since clouds are opaque to infrared radiation, there are many gaps in individual images, but this is partly overcome by producing composite maps over a few days. SST is also important operationally to provide a boundary condition for numerical [weather prediction models](#), and as a climate indicator. A time sequence of global SST maps from IR sensors has been available since 1981 (Kilpatrick et al., 2001). A recent major development in the field has been a move to create global SST maps, climatologies, and anomalies that combine the best data from several different types of sensors in a complementary approach (Donlon et al., 2007). One of the current challenges in the field is to safeguard the absolute accuracy and long-term stability of the

SST climate record even during adverse conditions of stratospheric aerosols following major volcanic eruptions.

The next class of sensors, *passive microwave radiometers*, detects the top of atmosphere brightness temperature in a number of different microwave bands and obtains additional independent information by separately measuring the H and V polarized radiation. Microwave emission from the sea is more complex than for infrared (Swift, 1980; Ulaby et al., 1981, 1982). While the blackbody emission is linear with absolute temperature, the emissivity is less than 0.5 and varies with the viewing angle (and hence the surface slope distribution over the field of view, typically driven by the local wind stress) and with the dielectric properties of seawater which depend on temperature and salinity. Consequently the actual dependence of the brightness temperature on the SST is very complex and requires information about the surface roughness and dielectric properties. The positive outcome of this complex behavior is that the additional dependencies provide the opportunity to retrieve wind speed and direction and salinity as well as temperature from microwave radiometry.

Microwaves pass through clouds but are still affected by water vapor and other gases. In practice the atmospheric correction to remove these effects and the geophysical algorithm to convert brightness temperature into the required ocean variable of SST are performed in a single empirical algorithm which is based on the different sensitivities of each different frequency and polarization channel to different factors (Wentz, 1997). Thus depending on which frequency bands they use, microwave radiometers deliver products of SST (based on the temperature of the top few mm of the sea), wind speed (and in one case wind direction), atmospheric water vapor, cloud liquid water, and rain over the sea. These are all useful measurements for oceanographers. Two sensors are now orbiting that are able to measure sea surface salinity (SSS) through the low-frequency microwave emission at about 1.4 GHz. The NASA Aquarius sensor uses three non-scanning radiometers each with a field of view of about 100 km that are swept along the ocean surface by the satellite trajectory (Lagerloef et al., 2008). The ESA SMOS (Soil Moisture and Ocean Salinity) mission (Drinkwater et al., 2009a) uses 69 radiometers in a “Y” formation to measure the microwave emission and an interferometric approach to aperture synthesis to provide a cross-track imaging capability with a retrieval cell size of 35–50 km (Drinkwater et al., 2009b).

Present microwave radiometers have a spatial resolution coarser than 50 km for SST, although they oversample to deliver pixels with a dimension of 0.25° . Since 2002, daily global SST maps have been routinely produced from microwave radiometry from the previous three days. Because they are unobstructed by cloud and return null data only where there is very heavy rain, they have a number of important applications in monitoring larger ocean features and serve as a complement to infrared-derived SST in operational contexts and in the

compilation of climate datasets (Chelton and Wentz, 2005; Donlon et al., 2007).

In contrast to the passive radiometers mentioned so far, active microwave devices supply their own energy, in the form of radar pulses which are emitted from the spacecraft, reflected from the sea surface, and received back at the sensor again. A major advantage is that the echo pulse can be compared to the emitted pulse in order to detect changes in its amplitude, shape, phase, or frequency, each of which may be used to derive particular information about the sea surface from which it was reflected. The effect of the intervening atmosphere can also be largely, though not entirely, ignored. The first class to consider is oblique viewing radars, and there are two special types used for ocean remote sensing, scatterometers, and synthetic aperture radars, normally referred to as SARs.

In the case of oblique viewing radars, when the emitted pulse encounters a smooth sea surface, most of its energy is reflected by specular reflection away from the sensor. If the surface is roughened, the incident energy is scattered more diffusely and with increasing roughness more energy is backscattered to the radar. While it is difficult to describe this behavior precisely (Ulaby et al., 1982), the main mechanism for backscatter at incidence angles between 20° and 70° is based on Bragg scattering theory. This shows that it is the elements of sea surface roughness with a wavelength of $\lambda_{sea} = \lambda_{radar} / 2 \sin \theta$ (where θ is the radar viewing incidence angle) that contribute most to the backscattering. The primary measured quantity is the radar backscatter cross section. This is the ratio between the returned and emitted signal, normalized for geometrical spreading; so, it changes only with the change in surface roughness and is usually given the symbol σ_0 . Depending on the radar frequency and its viewing angle, the Bragg wavelength is typically in the range between 1 and 30 cm. These are the small ripples that are produced by wind stress on the sea surface, and so oblique viewing radars are very sensitive to wind conditions over the sea.

A *scatterometer* exploits this in order to serve as a wind-monitoring instrument. It measures the average backscatter from a fairly wide region. The backscatter is compared with an empirical database relating σ_0 to wind speed and direction (Stoffelen and Anderson, 1997). A scatterometer is designed so that the same patch of sea is viewed from more than one direction within a few minutes, which enables both wind speed and direction to be retrieved, with a typical spatial resolution of 50 km. Using a single scatterometer the distribution of wind speed and direction can be sampled over the world ocean in about 2 days.

A SAR samples the return pulse in very fine detail and applies signal processing techniques to map at a spatial resolution of about 20 m, using the echoes from several hundred pulses to contribute to defining the magnitude of σ_0 at each pixel on the reconstructed image. The magnitude of σ_0 across a SAR image is set mainly by the spatial distribution of wind speed and direction, so a SAR image can map small-scale variability of the

wind field. However, local convergence, divergence, or shear in the sea surface currents causes modulation of the wind-driven Bragg ripples. This allows signatures of the underlying small-scale ocean dynamics to be created as patterns on the image. A variety of features have been revealed by this means, most notably internal waves, surface swell waves, bathymetric features in shallow tidal seas, ocean fronts, and small-scale eddies (Alpers and Hennings, 1984; Alpers, 1985; Robinson, 2004, pp. 521–555). Slicks, where surface-active material damps the amplitude of the Bragg waves, are also visible as dark patches on SAR images (Alpers and Hühnerfuss, 1988).

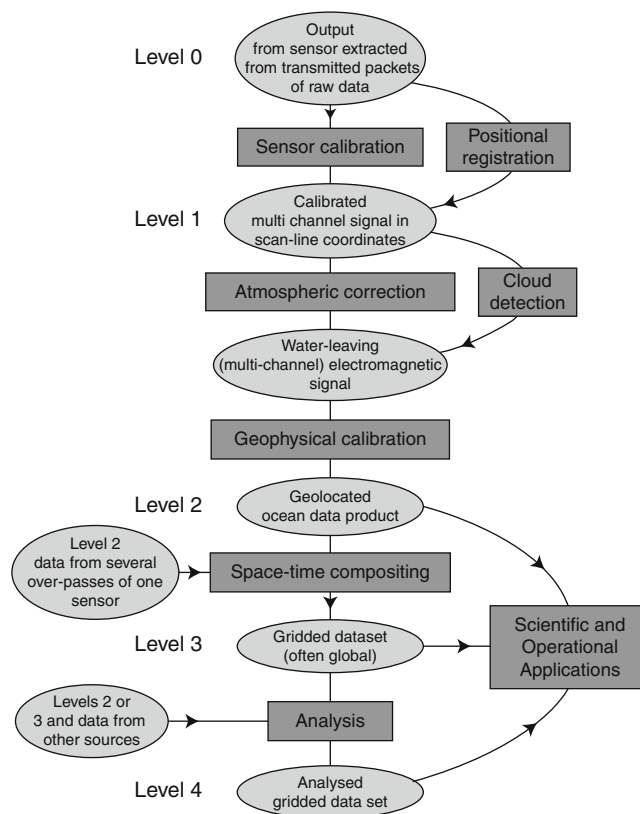
The final class of radars is the altimeter which views the surface at normal incidence using a nadir-pointing antenna. Its primary purpose is to measure the travel time of a radar pulse reflected from the sea surface and hence the distance between the altimeter and the sea surface. Precise knowledge of the satellite's position relative to the center of the Earth then allows the absolute height of the sea surface to be determined. As it travels along its orbit above the Earth, an altimeter can therefore measure the gradual along-track slope of the sea surface, which is another of the primary observable quantities of satellite oceanography. The slope of the sea surface over a distance of tens to hundreds of kilometers is an extremely important measurement since from it can be derived the surface geostrophic flow of the ocean. After corrections for the effect of the atmosphere on the speed of EMR, the most successful altimeters can measure height with a resolution of about 2 cm and therefore resolve much subtle detail in the time varying patterns of ocean circulation (Chelton et al., 2001). To make absolute retrievals of the geostrophic currents requires a determination of the geoid independent of altimetry measurements. Without this it is possible to measure only the variable part of the ocean dynamic surface topography, relative to the long-term mean altimeter height, which depends on both the geoid and the steady ocean surface current distribution that remains undetermined. Nonetheless, the availability of high accuracy altimetry since 1992 has greatly increased knowledge and understanding of the time variable component of the world ocean circulation and has had a great impact on dynamical oceanography. The challenge remains to be able to determine the absolute dynamic topography to an accuracy of 1 cm over horizontal distances less than 100 km, which requires knowledge of the geoid to the same accuracy. The European Gravity and Ocean Circulation Experiment (GOCE) was designed to meet this need (Johannessen et al., 2003; Bingham et al., 2011) as the best independent estimates of the geoid from the GRACE satellite gravity mission do not match this requirement, although by combining them with best estimates of the ocean mean circulation derived from hydrographic data, the absolute dynamic topography produced from altimeters has acceptable accuracy for many applications at lengths scales of 1,000 km and greater (Rio and Hernandez, 2004).

In addition to the altimeter's height measurement, the difference between the shape of the outward and return pulses is interpreted to provide an estimate of the sea surface wave height to an accuracy as good as that of wave buoys. The regular global mapping of significant wave height (SWH) for more than 20 years has greatly enhanced our knowledge of global wave statistics. Finally, the magnitude of the reflected energy is inversely related to the short-scale surface roughness from which the wind speed, but not direction, can be estimated. These capabilities make the satellite radar altimeter a highly versatile instrument, since each type of measurement is essentially independent of the other.

Producing ocean data products from raw satellite measurements

In the early years of Earth observation from satellites (the 1970s and 1980s), the motivation for new sensors and missions was primarily to develop the technological capabilities to prove the validity and utility of the satellite measurements to the communities who could make use of them. Users typically had to perform atmospheric corrections themselves and to develop the “geophysical algorithms” for deriving practically useful ocean quantities from the sensor's primary measurement. By the 1990s, the data processing infrastructure had evolved as the best processing procedures were validated and standardized. It is now a general practice for the agency managing a satellite to take responsibility for the whole chain of procedures needed to deliver calibrated and validated data products in the form of ocean quantities that can be applied immediately by users without specialist technical knowledge of the processing techniques. In the last decade the Internet has made it possible to broadcast data products widely within hours or less of the raw data acquisition.

Given the difference between the diverse methodologies identified in the previous section, the details of the processing chain to convert raw data into a useful data product differ considerably from sensor to sensor. Nonetheless, there is a generic set of tasks that need to be performed on data as part of the so-called ground segment of each Earth observation satellite mission, as illustrated in Figure 3. This is largely self-explanatory, but it is important for users of satellite data to know that the “level” assigned to a data product relates to the stage of processing that has been reached. Table 1 presents the definitions set by the Committee on Earth Observation Satellites (CEOS). Users requiring data containing oceanographically useful products would not be advised to use anything below level 2. At this level the data will be geographically defined, but in the native sampling grid of the sensor (aligned along and across the satellite track rather than north–south), and be at the finest resolution possible. It will have had atmospheric corrections applied (if appropriate) and will consist of useful ocean variables. Level 3 products will be resampled onto a regular geographical grid and may well be composed from several overpasses



Ocean, Measurements and Applications, Figure 3 An outline of the procedures required to convert raw satellite data into ocean data products ready for use in ocean science and marine-related applications. The data product levels are defined in Table 1.

of the same sensor. Level 4 products result from the analysis of data from sources beyond a single satellite and may include in situ data and the products of models.

As operational applications of satellite data products have matured, it is desirable for derived ocean variables to be accompanied by error estimates or values that allow the quality of the information in each pixel to be compared with other pixels in the same image and with other image products. For example, the cloud detection, atmospheric correction, or geophysical calibration processes may reach a more confident result for one pixel than another, as a result of which the user may wish to attach more weight to the influence of one pixel than another in driving a model or making an operational decision. Such an approach has already been established for SST processing (Donlon et al., 2007) and is emerging for others. It makes it essential to perform independent validation of the satellite data products, using coincident in situ measurements where possible.

Limitations of ocean remote sensing methods

As satellite measurements of ocean variables like SST, sea surface height anomaly, SWH, wind speed, and

Ocean, Measurements and Applications, Table 1 Levels of data product indicating the degree of processing that has been applied

Product level	CEOS definition
Raw data	Data in their original packets, as received from a satellite
0	Reconstructed unprocessed instrument data at full space-time resolution with all available supplemental information to be used in subsequent processing appended.
1	Unpacked, reformatted level 0 data at full space-time resolution with all supplemental information to be used in subsequent processing appended. Optional radiometric and geometric correction applied to produce parameters in physical units. Data generally presented as full time/space resolution. A wide variety of sublevel products are possible.
2	Retrieved environmental variables at the same resolution and location as the level 1 source data.
3	Data or retrieved environmental variables which have been spatially and/or temporally resampled (i.e., derived from level 1 or 2 products). Such resampling may include averaging and compositing.
4	Model output or results from analyses of lower level data (i.e., variables that are not directly measured by the instruments, but are derived from these measurements).

chlorophyll are well accepted, by most oceanographers, the inherent limitations of the approach must not be overlooked. Satellite data should complement rather than replace conventional ocean observations. There are many measurements that cannot be made from satellites and others that are better made by in situ sensors, depending on the particular application.

Firstly, as [Figure 1](#) shows, we cannot always obtain satellite measurements when and where we wish, being limited by the coupled sampling characteristics of scanning sensors in satellite orbits. It is inappropriate to rely on satellite data for observations of small-scale processes in estuaries and close to coasts, since the timescales are generally too short to sample adequately at fine spatial resolution.

Secondly, the calibration and validation of satellite-derived observations of ocean quantities must be based on in situ measurements of the same quantities. Even when geophysical algorithms are well calibrated, in situ data coincident with the satellite overpass must continue to be acquired for validation of a complete satellite mission. Ideally, validation samples should be well spread geographically. Since satellite sensors are unavailable for inspection at the end of a mission, the quality of their data sets becomes suspect without a satisfactory validation program.

Thirdly, remote sensing has a limited repertoire of ocean variables that can be measured, even given the innovative techniques developed to infer useful properties

from the four primary measurements that can be summarized as color, brightness temperature, surface roughness, and surface height. Most derived measurements are of physical properties. Apart from chlorophyll concentration derived from color, remote sensing makes no other marine biological measurements (excluding satellite tracking of tagged marine animals). Neither can chemical samples be made, other than salinity. The contribution of satellite data to marine biogeochemistry comes through providing a space-time overview of the physical environmental context in which conventional chemical and biological observations are interpreted.

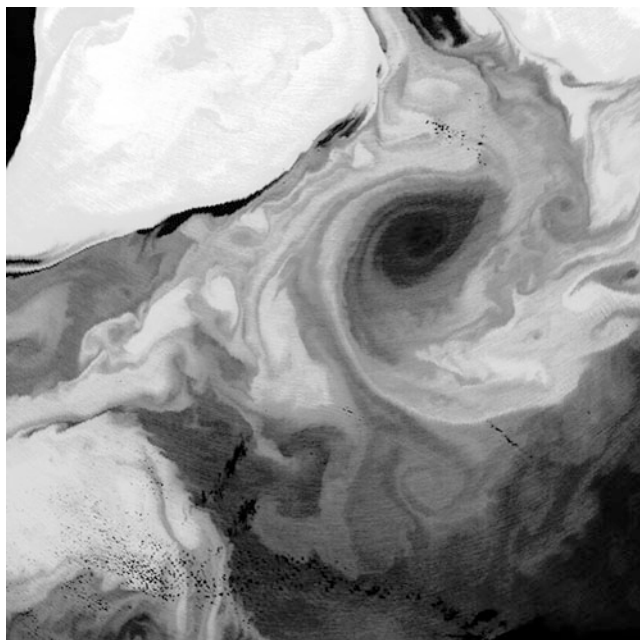
Finally, remote sensing relies on EMR emerging from the sea surface. For infrared and microwaves, this means that only quantities at the very surface of the sea can be detected. For visible wavebands, the light comes from the photic zone, typically 1 to 30 m deep depending on water turbidity. Present remote sensing methods do not measure any ocean properties below the ocean surface layer, which is a fundamental limitation that must always be taken into account when interpreting satellite data. That is not to say that remote sensing cannot be used to study phenomena occurring deeper in the water column, as explained below, but this is only possible if the phenomenon causes changes to the surface water that can be interpreted as a “surface signature” within satellite datasets.

Ocean applications of remote sensing

It misses the real strength of satellite data to consider them simply as an alternative to replace conventional in situ data. The unique contribution of satellite data to ocean science comes from spatial maps of ocean properties, repeated as a regular time series, that show how the detailed two-dimensional patterns of ocean surface properties vary with time. This is enhanced when different types of measurements from different classes of sensor are compared with each other, opening up new understanding of ocean processes. Therefore, the essence of satellite oceanography is to be found not simply in the technology and clever measurement methodologies, but in the creative ways in which satellite data are analyzed, blended with other observations, and assimilated in theoretical models, leading to new scientific insights about the way the ocean behaves.

Ocean phenomena observable from space

For the reasons implicit in [Figure 1](#), mesoscale variability is the class of ocean phenomena that are revealed particularly well in many types of satellite data. [Figure 4](#) demonstrates this with the [sea surface temperature](#) distribution in a 500 km square region of the S.W. Atlantic Ocean. Careful inspection shows a wealth of complex mesoscale eddy structures at a range of scales, from the central dominant eddy with a diameter of about 200 km down to small features of less than 10 km diameter. This “snapshot” of what is evidently a highly dynamic and evolving temperature



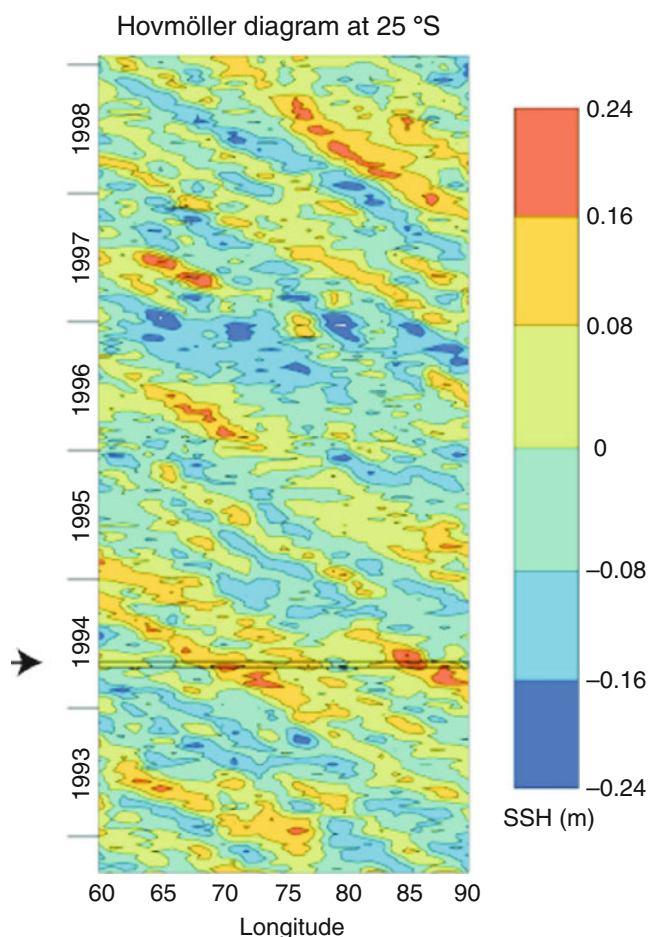
Ocean, Measurements and Applications, Figure 4 Sea surface temperature image of the S.W. Atlantic Ocean acquired on October 13, 1992, by the Along-Track Scanning Radiometer (ATSR) on the ERS-1 satellite of the European Space Agency. In this 500×500 km image with 1 km pixels, the contrast has been stretched to enhance the visibility of the mesoscale variability. The gray tone ranging from *black* to *white* represents a temperature range of 8°C , *black* being coldest and *white* being warmest.

distribution provides insight into the interrelationships and dynamic similarity between the scales. It provides a classic picture of two-dimensional turbulence in a rotating frame, at its true size and full complexity in place of the scaled-down laboratory experiments that had to be used before satellite images such as [Figure 4](#) became available. In cases where there is primary production occurring, similar mesoscale patterns can be seen in the chlorophyll maps derived from ocean color. The capacity to see the distribution of color and SST relative to each other is not only instructive for ocean dynamics but has led to a better understanding of the interaction between physical and biological processes at the mesoscale and given new insights to marine biologists (Saraceno et al., 2004, 2005). Satellite data confirm the presence of similar structured mesoscale features in all parts of the world ocean (Robinson, 2010, chs.3–5), although outside this very vigorous confluence of the Brazil and Malvinas currents, their signatures may be much weaker in terms of the temperature gradients across them. The trajectory of larger eddies can be tracked in the surface height maps from altimetry (Schonten et al., 2000). Coastal upwelling is also evident from the SST field as ribbons of cold water along coastlines and can be correlated with the wind forcing of upwelling, using satellite-derived wind field maps.

Exploring beyond individual features, the statistical characteristics of the variability field can also be monitored from space. From altimetry it is possible to map the eddy kinetic energy as a derived variable (Ducet et al., 2000). Using automatic detection of fronts on SST images, it is possible to analyze long records of satellite data to produce climatologies of fronts (Belkin and Cornillon, 2003). Fronts, upwelling, and mesoscale eddy activity, all promote vertical mixing and transport of nutrients to the surface, supporting enhanced primary production. Basin scale satellite maps of chlorophyll from ocean color can be compared with the mapped eddy activity, fronts, and upwelling to illuminate the ways in which ocean dynamics influence marine biology on the large scale. Fisheries science benefits from such an approach (Santos, 2000). Studies on this scale, with an ocean-wide perspective that looks at the integrated effect of many individual mesoscale features, are needed to understand the Earth system as a whole. They are dependent on satellite mapping of ocean properties integrated on the global scale, but which must be derived from fine-scale images to avoid aliasing.

Another class of ocean phenomenon revealed readily in satellite data are the large-scale ocean waves whose dynamics are associated with the planetary beta effect, a consequence of the local Earth rotation parameter varying with latitude over the Earth. [Figure 5](#) illustrates how such waves show up when a time series of satellite maps over an ocean basin are sliced to show how an ocean property such as sea surface height varies with longitude (x -axis) and time (y -axis) (Chelton and Schlax, 1996). The diagonally trending structures provide evidence of westward propagating features. In many cases these can be identified as Rossby waves which are very difficult to detect using conventional observations, mainly because their timescale is so large, taking years to cross an ocean basin. Their ready detection from satellites has led to a healthy interplay with theoretical modelers as their early theories were tested against the new observational evidence from satellites and found to be in need of refinement (Killworth and Blundell, 2003). Having used altimetry to confirm the existence of Rossby waves, it was interesting to discover that they also had a signature in the global maps of SST anomaly and even in ocean color (Cipollini et al., 2006), which prompted further theoretical analysis (Charria et al., 2006).

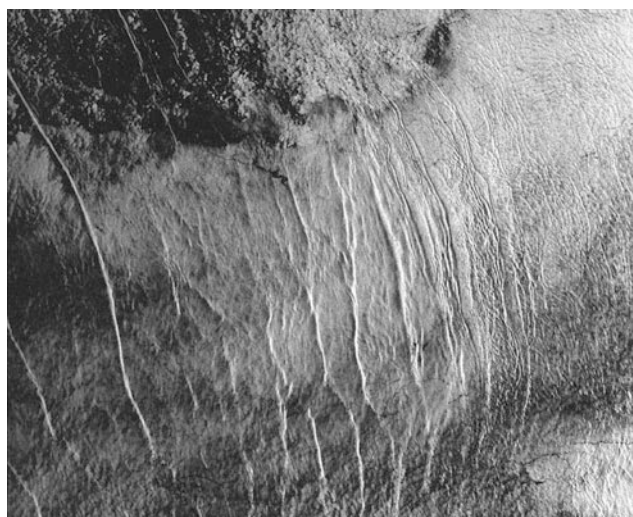
In general, remote sensing has proved in practice to be most readily accepted by oceanographers when it delivers unique data for displaying slow, large-scale processes. A good example is the El Niño phenomenon in the tropical Pacific Ocean (Picaut et al., 2002). Satellites deliver SST anomaly and surface height anomaly maps which show the strong characteristic differences between El Niño, non-El Niño, and La Niña conditions. They provide maps of wind speed and direction, and rainfall over the sea, which are relevant to the atmosphere–ocean coupling that is central to understanding how to model and predict the onset of an El Niño event. They also deliver maps of



Ocean, Measurements and Applications, Figure 5 A time versus longitude plot of sea surface height (SSH) measured by altimeter at 25 °S in the Indian Ocean between 60° and 90°E. The arrow and lines define a single contribution to the plot, from the SSH map for Cycle 60 (1–11 May 1994) of the TOPEX/POSEIDON mission (Figure courtesy of Paolo Cipollini).

chlorophyll concentration showing the impact on primary production and hence the fisheries resources. Another category of phenomena whose study has benefited from satellite data is tropical cyclones (called hurricanes when they occur in the Atlantic Ocean), their prediction, tracking, and the changes they leave behind in the ocean. Passive microwave radiometers have the unique capacity to monitor SST under large amounts of cloud surrounding hurricanes. A third example concerns the unprecedented melting of the Arctic sea ice during summer 2007 and again in 2012, which satellites monitored as it happened. Once the ice had melted, the temperature on the newly exposed sea could be monitored until the surface froze over again.

In addition to these meso- and large-scale features that lend themselves to observation from space, there is a variety of other smaller scale phenomena in coastal and



Ocean, Measurements and Applications, Figure 6 SAR image showing internal solitons propagating from left to right across the Andaman Sea. This image from the ERS-1 SAR is 100 km across (see Robinson, 2004, Figure 10.47).

shelf seas that have benefited from satellite data. Where shallow seas are well mixed vertically, then the surface properties measured by remote sensing are uniform throughout the water column. Here SST characterizes the total heat content of the sea, and ocean color represents the whole depth, allowing chlorophyll concentration or the suspended particulate content to be measured. In summer, when the deeper parts of tidal shelf seas start to stratify, the location of the frontal edge shows up as a characteristic signature on SST images. Studies in this field were among the earliest cases where satellite data made a significant contribution to oceanographic research (Simpson and Bowers, 1979).

Another class of phenomena revealed very clearly on SAR images when the wind conditions are appropriate is that of internal waves which propagate on the thermocline (Alpers, 1985). This is a clear example of how remote sensing can provide important information from a unique view of a subsurface ocean process, centered as much as 50 m below the surface. Through the action of the convergence of currents associated with internal waves modulating the sea surface roughness, the phase pattern (troughs and crests) of the internal waves is “painted” in the SAR images of σ_0 . Thus it is possible to differentiate between different wave trains and how they refract as they propagate across a region. Figure 6 illustrates a particular type of large-amplitude nonlinear internal wave, an internal soliton (Liu et al., 1998), in this case propagating across the Andaman Sea toward the Thailand coast. Even though such features have only a very small amplitude at the sea surface, the thermocline may rise and fall by up to 50 m as the soliton passes and so the ability to monitor them from satellites is very valuable.

Operational applications of satellite oceanography

In the last two decades, operational applications have followed in the wake of mature scientific understanding of what can be learned about the ocean from satellite data. The simplest and most direct of these are based on providing satellite images to mariners, as in the case of SST and surface chlorophyll maps being used by deep sea fishing fleets. In coastal regions where fish farming and other aquaculture activities are threatened by harmful algae, remote sensing is used to provide advance warning of phytoplankton blooms, although ship sampling is needed to determine whether the algal species is problematic. Wave height data from altimetry are made publicly available in near-real time (NRT – within minutes to hours of acquisition) so that mariners can keep a watch on developing sea state and validate forecasts. Rather more elaborate, but still basically image based, is the use of SAR data for surveillance of oil spills, both for enforcement of antipollution legislation and to alert pollution protection agencies. For example, the European Marine Safety Agency has developed the CleanSeaNet service, which provides for the European seas a range of detailed information to Member States, including rapid delivery of available satellite images and oil slick position (EMSA, 2008).

Another important operational use of satellite ocean data is the incorporation of wind measurements from scatterometers into numerical [weather forecasting](#). Although this might be considered to be a purely meteorological application, the benefits to the marine community in terms of improved offshore weather forecasts are significant. Ocean wave forecasting is closely linked to weather forecasting, being based on wind forcing of ocean models, but increasing use is now being made of direct wave measurements (SWH from altimetry or wave spectra from SAR) for assimilation into, or validation of, third-generation wave models.

A much more complex, interactive, and far-reaching use of satellite ocean data for ocean forecasting systems is in the process of becoming operationally active in Europe (Johannessen et al., 2006). This is where NRT measurements of SST and surface height anomaly derived from satellite data are assimilated into forecasting systems based on numerical ocean general circulation models. These generate nowcasts and short-term forecasts of the physical properties of the ocean, at depth as well as at the surface. NRT observations from subsurface floats and satellites are needed to constrain the models so that they follow the actual ocean behavior. Meeting the model requirements of resolution, timeliness, and accuracy has presented a challenge to satellite ocean data producers, which has been met through the complementary use of products from different satellites in both the SST (Donlon et al., 2007) and surface height (Pascual et al., 2006) communities. It is intended that in future the ocean models will include biogeochemical components forecasting primary production and using products derived from satellite

ocean color data as assimilation variables. Finally, satellite data are recognized to have an essential contribution to make in monitoring changes in the climate of the ocean in response to global warming.

Conclusions

There is little doubt that the increasing availability of well-calibrated ocean measurements derived from satellites has changed the way oceanographers go about their science. It has encouraged both a more global approach with an emphasis on understanding the spatial structures of phenomena and a more holistic approach, allowing new views of the connections between marine biology, dynamics, physics, and biogeochemical processes. Moreover, the disciplined delivery of data within the operational constraints of timeliness and accuracy has increased the wider impact of satellite oceanography on many aspects of human interaction with the ocean for transport, food, industry, and leisure.

There remain significant technical challenges that include the acquisition of a more precise ocean geoid, beyond that provided by GOCE, improving the accuracy and resolution of salinity measurements from space and improving techniques for directly measuring surface currents at short-length scales from interferometric radars (Romeiser and Runge, 2007) or from Doppler analysis of SAR data (Chapron et al., 2005). There is also a political challenge to develop and maintain the continuity of existing satellite ocean observing capacity within an international infrastructure of open access to data. This will be essential to provide us with knowledge of how the ocean changes in coming decades in response to the accelerating impact of man's occupation of the planet, and it will hopefully contribute to an internationally shared strategy to reduce that impact before its effect has grown beyond control.

Bibliography

- Alpers, W., 1985. Theory of radar imaging of internal waves. *Nature*, **314**, 245–247.
- Alpers, W., and Hennings, I., 1984. A theory of the imaging mechanism of underwater bottom topography by real and synthetic aperture radar. *Journal of Geophysical Research*, **89**, 10529–10546.
- Alpers, W., and Huhnerfuss, H., 1988. Radar signatures of oil films floating on the sea surface and the Marangoni effect. *Journal of Geophysical Research*, **93**, 3642–3648.
- Belkin, I. M., and Cornillon, P., 2003. SST fronts of the Pacific coastal and marginal seas. *Pacific Ocean and Pacific Oceanography*, **1**, 90–100.
- Berger, M., et al., 2002. Measuring ocean salinity with ESA's SMOS mission. *ESA Bulletin*, **111**, 113–121.
- Bingham, R. J., et al., 2011. An initial estimate of the North Atlantic steady-state geostrophic circulation from GOCE. *Geophysical Research Letters*, **38**, L01606.
- Brown, S. J., et al., 1997. New aerosol-robust sea surface temperature algorithms for the along-track scanning radiometer. *Journal of Geophysical Research*, **102**, 27973–27990.

- Chapron, B., et al., 2005. Direct measurements of ocean surface velocity from space: interpretation and validation. *Journal of Geophysical Research*, **110**. doi:10.1029/2004JC002809.
- Charria, G., et al., 2006. Understanding the influence of Rossby waves on surface chlorophyll concentrations in the North Atlantic Ocean. *Journal of Marine Research*, **64**, 43–71.
- Chelton, D. B., and Schlax, M. G., 1996. Global observations of oceanic Rossby waves. *Science*, **272**, 234–238.
- Chelton, D., and Wentz, F. J., 2005. Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research. *Bulletin of the American Meteorological Society*, **86**(8), 1097–1115.
- Chelton, D. B., et al., 2001. Satellite altimetry. In Fu, L.-L., and Cazenave, A. (eds.), *Satellite Altimetry and Earth Sciences*. San Diego, CA: Academic Press, pp. 1–131.
- Cipollini, P., et al., 2006. Remote sensing of extra-equatorial planetary waves. In Gower, J. F. R. (ed.), *Manual of Remote Sensing, Volume 6: Remote Sensing of Marine Environment*. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, pp. 61–84.
- Cracknell, A. P., 1997. *The Advanced Very High Resolution Radiometer (AVHRR)*. London: Taylor and Francis.
- Donlon, C. J., et al., 2007. The global ocean data assimilation experiment (GODAE) high resolution sea surface temperature pilot project (GHRSSST-PP). *Bulletin of the American Meteorological Society*, **88**. doi:10.1175/BAMS-88-8-1197.
- Drinkwater, M., et al., 2009a. Exploring the water cycle of the ‘blue planet’ – the soil moisture and ocean salinity (SMOS) mission. *ESA Bulletin*, **137**, 7–15.
- Drinkwater, M., et al., 2009b. Star in the sky – The SMOS payload: MIRAS. *ESA Bulletin*, **137**, 17–22.
- Ducet, N., et al., 2000. Global high-resolution mapping of ocean circulation from the combination of T/P and ERS-1/2. *Journal of Geophysical Research*, **105**, 19477–19498.
- EMSA, 2008. “CleanSeaNet”. Retrieved 14 July 2008, from <http://cleanseanet.emsa.europa.eu/index.html>.
- Ewing, G. C. (ed.), 1965. *Oceanography from Space*. Ref. 65–10. Woods Hole, MA: Woods Hole Oceanographic Institution.
- Fu, L.-L., and Cazenave, A. (eds.), 2001. *Satellite Altimetry and Earth Sciences*. San Diego: Academic.
- Gade, M. (ed.), 2005. *Surface Slicks and Remote Sensing of Air-Sea Interaction*. Kluwer.
- Gordon, H. R., 1997. Atmospheric correction of ocean color imagery in the Earth Observing System era. *Journal of Geophysical Research*, **102**, 17081–17106.
- Gower, J. F. R. (ed.), 1981. *Oceanography from Space*. New York/London: Plenum.
- Halpern, D. (ed.), 2000. *Satellites, Oceanography and Society*. Elsevier: Amsterdam, 380 pp.
- Johannessen, J., et al., 2003. The European gravity field and steady-state ocean circulation explorer satellite mission: impact in geophysics. *Surveys in Geophysics*, **24**, 339–386.
- Johannessen, J. A., et al., 2006. Marine environment and security for the European area: toward operational oceanography. *Bulletin of the American Meteorological Society*, **87**, 1081–1090.
- Killworth, P. D., and Blundell, J. R., 2003. Long extra-tropical planetary wave propagation in the presence of slowly varying mean flow and bottom topography II: ray propagation and comparison with observations. *Journal of Physical Oceanography*, **33**, 802–821.
- Kilpatrick, K. A., et al., 2001. Overview of the NOAA/NASA advanced very high resolution Pathfinder algorithm for sea surface temperature and associated matchup database. *Journal of Geophysical Research*, **106**, 9179–9197.
- Lagerloef, G., et al., 2008. The Aquarius/SAC-D Mission: designed to meet the salinity remote-sensing challenge. *Oceanography*, **21**, 68–81.
- Liu, A. K., et al., 1998. Evolution of nonlinear internal waves in the East and South China Seas. *Journal of Geophysical Research*, **103**, 7995–8008.
- Martin, S., 2004. *An Introduction to Ocean Remote Sensing*. Cambridge: Cambridge University Press.
- Maul, G., 1985. *Introduction to Satellite Oceanography*. Dordrecht: Martinus Nijhoff.
- McClain, C., et al., 2006. Satellite data for ocean biology, biogeochemistry, and climate research. *EOS Transactions*, **87**, 337–343.
- McMillin, L. M., and Crosby, D. S., 1984. Theory and validation of the multiple window sea surface temperature technique. *Journal of Geophysical Research*, **89**, 3655–3661.
- O’Reilly, J. E., et al., 1998. Ocean colour chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, **103**, 24937–24953.
- Pascual, A., et al., 2006. Improved description of the ocean meso-scale variability by combining four satellite altimeters. *Geophysical Research Letters*, **33**. doi:10.1029/2005GL024633.
- Picaut, J., et al., 2002. Mechanisms of the 1997–1998 El Niño-La Niña, as inferred from space-based observations. *Journal of Geophysical Research*, **107**. doi:10.1029/2001JC000850.
- Rio, M.-H., and Hernandez, F., 2004. A mean dynamic topography computed over the world ocean from altimetry, in situ measurements, and a geoid model. *Journal of Geophysical Research*, **109**. doi:10.1029/2003JC002226.
- Robinson, I. S., 1985. *Satellite Oceanography: An Introduction for Oceanographers and Remote Sensing Scientists*. Chichester: Ellis Horwood.
- Robinson, I. S., 2004. *Measuring the Ocean from Space: The Principles and Methods of Satellite Oceanography*. Chichester: Praxis Publishing with Springer.
- Robinson, I. S., 2010. *Discovering the Ocean from Space: The Unique Applications of Satellite Oceanography*. Heidelberg: Springer.
- Romeiser, R., and Runge, H., 2007. Detailed analysis of ocean current measuring capabilities of TerraSAR-X in several possible along-track InSAR modes on the basis of numerical simulations. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 21–35.
- Saltzman, B. (ed.), 1985. *Satellite Ocean Remote Sensing. Advances in Geophysics*. Orlando, FL: Academic Press.
- Santos, A. M. P., 2000. Fisheries oceanography using satellite and airborne remote sensing methods: a review. *Fisheries Research*, **49**, 1–20.
- Saraceno, M., et al., 2004. Brazil Malvinas Frontal System as seen from 9 years of advanced very high resolution radiometer data. *Journal of Geophysical Research*, **109**. doi:10.1029/2003JC002127.
- Saraceno, M., et al., 2005. On the relationship between satellite-retrieved surface temperature fronts and chlorophyll a in the western South Atlantic. *Journal of Geophysical Research*, **110**. doi:10.1029/2004JC002736.
- Schonten, M. W., et al., 2000. Translation, decay and splitting of Agulhas rings in the southeastern Atlantic Ocean. *Journal of Geophysical Research*, **105**, 21913–21925.
- Simpson, J. H., and Bowers, D. G., 1979. Shelf sea fronts’ adjustments revealed by satellite IR imagery. *Nature*, **280**, 648–651.
- Stewart, R., 1985. *Methods of Satellite Oceanography*. San Diego, CA: University of California Press.
- Stoffelen, A. C. M., and Anderson, D. L. T., 1997. Scatterometer data interpretation: estimation and validation of the transfer function CMOD4. *Journal of Geophysical Research*, **102**, 5767–5780.
- Swift, C. T., 1980. Passive microwave remote sensing. *Boundary Layer Meteorology*, **18**, 25–54.

- Ulaby, F. T., et al., 1981. *Microwave Remote Sensing: Active and Passive. I – Microwave Remote Sensing Fundamentals and Radiometry*. Reading, MA: Addison-Wesley.
- Ulaby, F. T., et al., 1982. *Microwave Remote Sensing: Active and Passive. II – Radar Remote Sensing and Surface Scattering and Emission Theory*. Reading, MA: Addison-Wesley.
- Wentz, F. J., 1997. A well calibrated ocean algorithm for SSM/I. *Journal of Geophysical Research*, **102**, 8703–8718.

Cross-references

Coastal Ecosystems
 Fisheries
 Global Climate Observing System
 Ocean Data Telemetry
 Ocean Internal Waves
 Ocean Measurements and Applications, Ocean Color
 Ocean Modeling and Data Assimilation
 Sea Ice Concentration and Extent
 Sea Surface Temperature
 Sea Surface Salinity
 Sea Surface Wind/Stress Vector

OCEAN-ATMOSPHERE WATER FLUX AND EVAPORATION

W. Timothy Liu and Xiaosu Xie
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Definition

The ocean–atmosphere water exchange is the difference between evaporation and precipitation at the surface of the ocean. Evaporation is the turbulent transport of water vapor from ocean to the atmosphere. Precipitation is the return of water to the ocean from the atmosphere in the form of rain and snow.

Introduction

The equation of water conservation in the atmospheric column is

$$\frac{\partial W}{\partial t} + \nabla \cdot \Theta = E - P = F \quad (1)$$

where

$$\Theta = \int_0^{p_s} q u dp \quad (2)$$

is the moisture transport integrated over the depth of the atmosphere, and

$$W = \frac{1}{g} \int_0^{p_s} q dp \quad (3)$$

is the precipitable water or column-integrated water vapor. In these equations, p is the pressure, p_s is the pressure at the surface, and q and u are the specific humidity and wind

vector at a certain level. Bold symbols represent vector quantities. F is the freshwater exchange between the ocean and the atmosphere and is the difference between evaporation (E) and precipitation (P) at the surface. The first term is the change of storage. For periods longer than a few days, it is negligible, and there is a balance between the divergence of the transport ($\nabla \cdot \Theta$) and the surface flux. The balance gives rise to two ways of estimating the fresh water flux. One is to measure E and P separately; the other is to estimate Θ .

The first method, through the small turbulent-scale processes, has been called the “supply side” estimation; the water is supplied by transport from the ocean. The second method has been called the “demand side” estimation; the large-scale atmospheric circulation demands the water flux from the ocean (WCRP, 1983). One of the most advanced statistical techniques, support vector regression (SVR), has been used to retrieve surface specific humidity (q), E , and Θ , from space-based data.

The scientific need of the flux is presented in section [Significance](#). Space-based estimations of q and E are described in section [Bulk Parameterization: The Supply Side](#). The validation of the water flux is difficult because of the lack of credible direct measurements. The conservation principles post constraints of the accuracy of these fluxes. The space-based estimation of $\nabla \cdot \Theta$ as water flux is described in section [Divergence of Moisture Transport: The Demand Side](#), which includes the validation through mass conservation of global ocean and the continent of South America. In turn, $\nabla \cdot \Theta$ is used as a constraint to the accuracy of E retrieval ([Equation 1](#)) in section [Marine Atmosphere Water Conservation](#). The feasibility of applying further constraints is explored in section [Ocean Heat and Surface Salinity Conservation](#).

There are many programs to produce P (e.g., Huffman et al., 1997; Adler et al., 2003; Joyce et al., 2004). The Tropical Rainfall Measuring Mission (TRMM, Kummerow et al., 2000) measures rainfall between 38° latitude north and south of the equator and has provided important calibration of P since 1998. The Global Precipitation Mission (GPM) will extend the coverage to extratropical regions, with increased sensitivity and accuracy. See Rainfall, by R. Ferraro, in this book for further discussion.

Significance

Water is the essential element for life. Over 70 % of the Earth’s surface is covered by the ocean, which forms the largest reservoir of water on Earth. The never-ending recycling process in which a small fraction of water is continuously removed from the ocean as excess evaporation over precipitation into the atmosphere, redistributed through atmospheric circulation, deposited as excess precipitation over evaporation on land, and returned to the ocean as river discharge, is critical to the existence of human life and the variability of weather and climate.

With their high specific heat and large thermal inertia, the oceans are also the largest reservoir of heat and the fly-wheel of the global heat engine. Since water has high latent heat, evaporation is also an efficient way to transfer the energy. Besides releasing latent heat to the atmosphere, the water evaporated from the surface forms clouds, which absorb and reflect radiation. Water vapor is also an important greenhouse gas, which absorbs more long-wave radiation emitted by Earth than the short-wave radiation from the Sun. Redistribution of clouds and water vapor changes the Earth's radiation balance.

The hypothesis of the amplification of water cycle resulted from global warming, which essentially states that wet places get wetter and dry places get dryer, is a typical problem joining the water and energy balances. Increase in global mean precipitation down to the surface has to be balanced by equal amount of evaporation from the surface to conserve water in the atmosphere. Increase in latent heat from the surface carried by evaporation requires increase in long-wave radiation down to the surface; any imbalance will result in climate changes.

The differential heating of the atmosphere by the ocean fuels atmospheric circulation, which in turn drives ocean currents. Both wind and current transport and redistribute heat and greenhouse gases. Adding heat and water changes density of air and seawater. The heat and water fluxes, therefore, change both the baroclinicity and stability (horizontal and vertical density gradients) of the atmosphere and the ocean. These in turn modify the shears of wind and current.

Bulk parameterization: the supply side

Most productions of space-based evaporation data sets in the past were based on bulk parameterization. Latent heat flux (LH) is related to E by the nearly constant value of latent heat of vaporization (L): $LH = L \times E$. LH , rather than E , is used in many of the past studies. The two parameters are used interchangeably in this entry, and our discussion on E applies equally to LH .

The computation of E by the bulk parameterization requires sea surface temperature (SST), wind speed (u), and q .

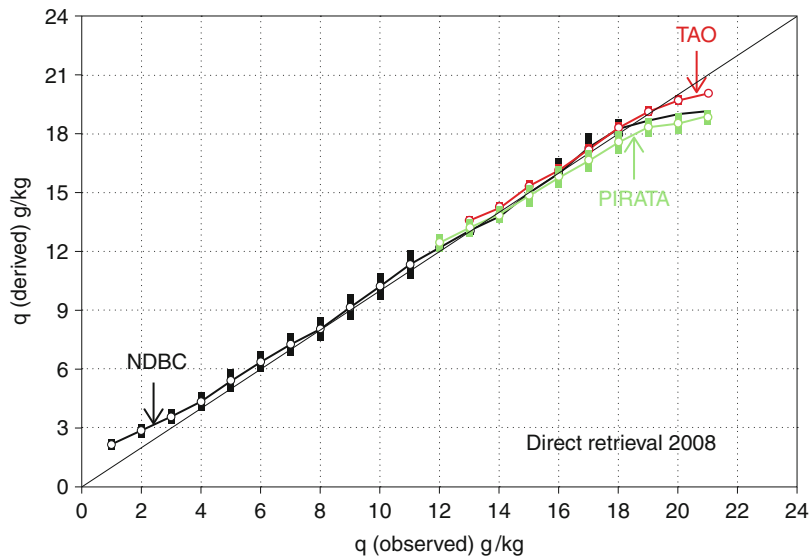
$$E = C_E \rho u (q_s - q) \quad (4)$$

where C_E is the transfer coefficient and ρ is the surface air density. q_s is usually taken to be the saturation humidity at SST multiplied by a factor of 0.98 to account for the effect of salt in the water. u and q should be measured in the atmospheric surface (constant flux) layer, usually taken at a reference level of 10 m. Over the ocean, u and SST have been measured from space, but not q . A method of estimating E using satellite data was demonstrated by Liu and Niiler (1984), based on an empirical relation between W and q on a monthly timescale over the global ocean (Liu, 1986). The physical rationale is that the vertical distribution of water vapor through the whole depth of the atmosphere is coherent for periods longer than

a week (Liu et al., 1991). The relation has been scrutinized in a number of studies and many variations of this method have been proposed to improve on the estimation (see Liu and Katsaros, 2001, for a review of earlier studies). Modification of this method by including additional estimators has been proposed (e.g., Wagner et al., 1990; Cresswell et al., 1991; Miller and Katsaros, 1991; Chou et al., 1995) with various degrees of improvement. Recently, neural network has also been used to mitigate the nonlinearity problem in deriving q (Jones et al., 1999; Bourras et al., 2002; Roberts et al., 2010). Algorithms to retrieve q from brightness temperatures (BT) measured by microwave radiometers were developed and improvements were demonstrated (e.g., Schulz et al., 1997; Schlüssel et al., 1995; Jackson et al., 2009). Yu and Weller (2007) have combined space-based observations with model output. Figure 1 shows the validation of q derived from BT measured by the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) through a statistical model built on SVR. The model outputs are compared with coincident q measured at buoys. For the year of 2008, 30,000 buoy data were randomly selected for validation. The mean and root mean square (rms) differences are 0.05 and 1.05 g/kg, respectively. The rms difference is only 5 % of the range of 20 g/kg and the statistical model appears to be successful. However, E depends on $\Delta q = q_s - q$, which is the small difference between the two large terms (q_s and q), and a small percentage error in q may still cause a large error in Δq and E .

Liu (1990) suggested and demonstrated two potential ways to improve E retrieval from satellite data. The first is to incorporate information on vertical distribution of humidity given by atmospheric sounders. Jackson et al. (2009) have recently adopted this suggestion. The other is to retrieve E directly from the radiances, since all the bulk parameters used in the traditional method could be derived from radiances measured by a microwave radiometer. The direct retrieval method may improve accuracy in two ways. The first is the use of a single CE to derive E (in training the statistical model). The second is to mitigate the magnification of error caused by multiplying inaccurate measurement of wind speed with inaccurate measurements of humidity (q and q_s) in the bulk formula.

Figure 2 compares the uncertainties of two sets of LH derived from the two methods. For the first set, SST and u from AMSR-E produced by Remote Sensing System (Wentz and Meissner, 2000) are used with q derived from AMSR-E BT (same as those in Figure 1). The second set is the output of a statistical model built on SVR, predicting E from the 12-channel AMSR-E BT. A total of 30,000 randomly selected LH computed from three groups of buoy data in 2008 are used in the validation exercise. Direct retrieval of daily mean values reduces the rms difference from 77 W/m² of the bulk parameterization method to 38 W/m². This is equivalent to a reduction from 19 % to 9.7 % of the dynamic range of 400 W/m².



Ocean-Atmosphere Water Flux and Evaporation, Figure 1 Bin-average of near surface specific humidity (q) derived from the statistical model compared with values measured at three groups of buoys. Standard deviation is superimposed on each bin-average as error bar.

The available E (or LH) products and the bulk parameters used to derive them exhibit substantial differences even for monthly means (e.g., Brunke et al., 2002; Bourras, 2006; Smith et al., 2011; Santorelli et al., 2011). The results of our direct retrieval depend on the mix of training data (buoy and ship measurements and NWP products). The conservation principle Equation 1 and the demand-side evaluation may serve as an effective way to evaluate current E products.

Divergence of moisture transport: the demand side

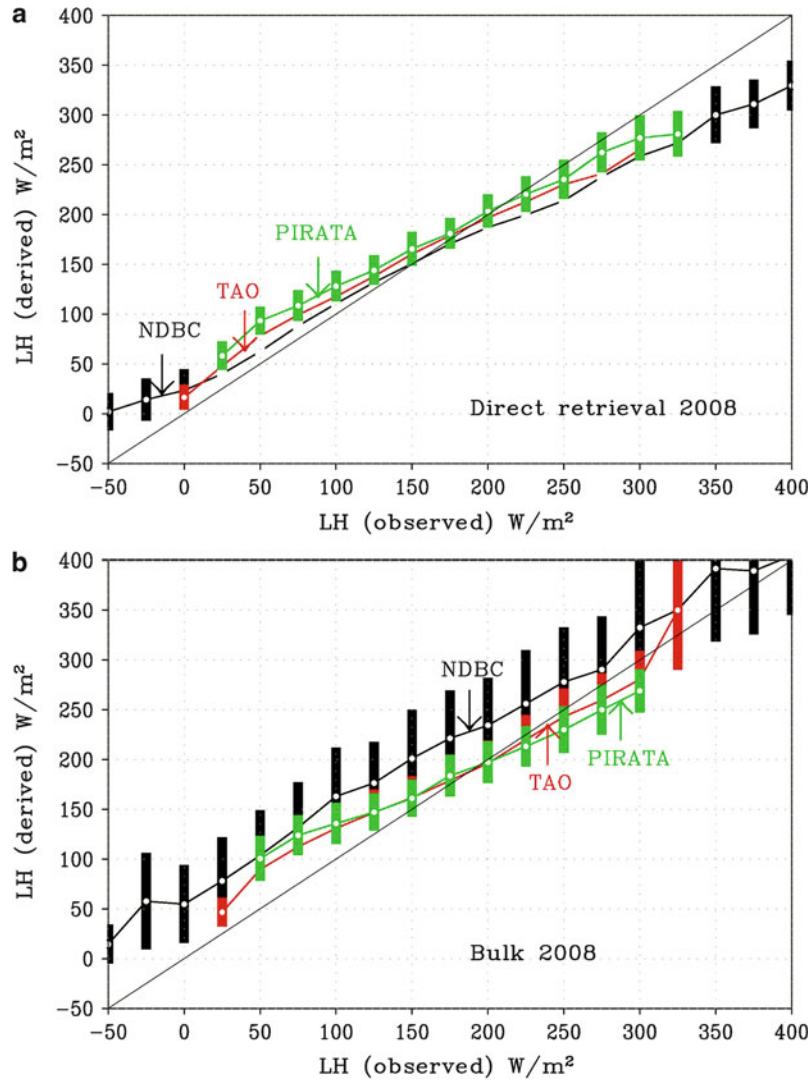
The computation of Θ , as defined in Equation 2, requires the vertical profile of q and u , which are not measured by space-based sensors with sufficient resolution. Θ can be viewed as the column of water vapor W advected by an effective velocity u_e , so that $u_e = \Theta/W$, and u_e is the depth-averaged wind velocity weighted by humidity. W has been derived from microwave radiometer measurements with good accuracy. Methods were developed to relate u_e to the equivalent neutral wind measured by scatterometers, u_s , based on polynomial regression (Liu, 1993) and neural network (Liu and Tang, 2005). Xie et al. (2008) added cloud-drift winds at 850 mb to u_s and used SVR instead of neural network. The scatterometer measurement and the cloud-drift winds represent ocean surface stress and free-stream velocity, respectively. Xie et al. (2008) showed that Θ derived from their statistical model agrees with Θ derived from 90 rawinsonde stations from synoptic to seasonal timescales and from equatorial to polar oceans. Hilburn (2010) found very good agreement between this data set and data computed from Modern Era Retrospective-analysis for Research and

Applications (MERRA) over the global ocean. MERRA is a NASA atmospheric reanalysis using a major new version of the Goddard Earth System Data Assimilation System (Rienecker et al., 2011). Figure 3 shows that, for a total of 26,000 pairs randomly selected data, 2/3 from rawinsonde and 1/3 from the reanalysis, the rms difference is 57.5 kg/m/s and the correlation coefficient is 0.95 for zonal component, and 49.7 kg/m/s and 0.89 for meridional component, for a range of approximately -600 to $+600$ kg/m/s.

Validation of our space-based estimation of $\nabla \cdot \Theta$ (as F) was achieved through mass balance of oceans and continents, using data of the Gravity Recovery and Climate Experiment (GRACE), which is a geodesy mission to measure Earth's gravity field. The variations of the gravity field are largely the results of the change of water storage. The air-sea water flux given by $\nabla \cdot \Theta$ integrated over all ocean area, together with river discharge (R) from all continents, should balance the rate of mass change ($\partial M/\partial t$) of all oceans:

$$\iint \frac{\partial M}{\partial t} = \int R - \iint \nabla \cdot \Theta \quad (5)$$

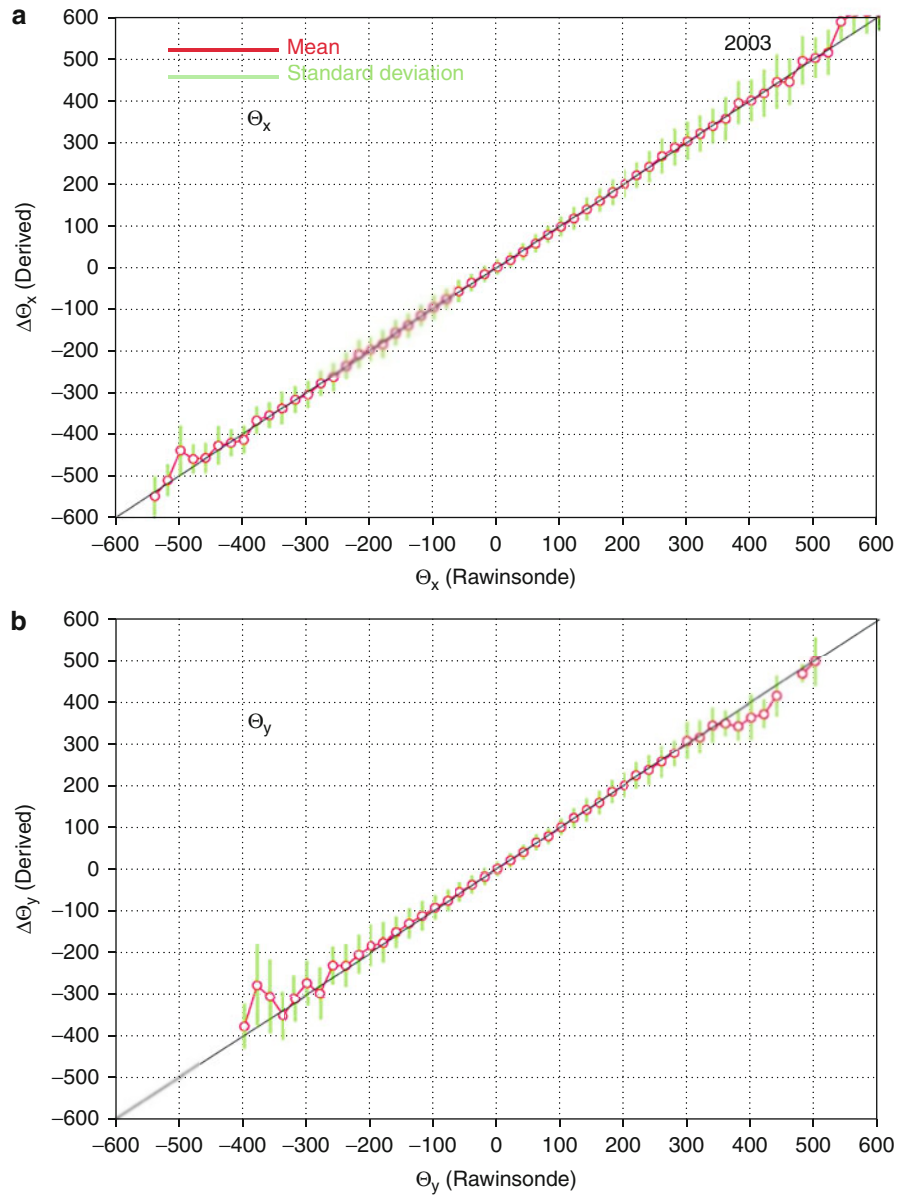
where $\int$ and $\iint$ represent line and area integrals, respectively. Figure 4 shows that monthly rate of mass change ($-\partial M/\partial t$), measured by GRACE, integrated over all oceans, agrees in magnitude and in phase with $\nabla \cdot \Theta$, derived from the statistical model of Xie et al. (2008) integrated over all ocean areas minus the line integral of R over all coastlines. The difference between $-\int \partial M/\partial t$ and $\iint \nabla \cdot \Theta - \int R$ has a mean of 2.1×10^8 kg/s and a standard deviation of 2.6×10^8 kg/s. The standard



Ocean-Atmosphere Water Flux and Evaporation, Figure 2 Bin-average of LH derived directly from the satellite measured radiances (a), and computed from bulk parameters (b), compared with coincident measurements at three groups of buoys. Standard deviation is superimposed on each bin-average as error bar.

deviation is 18 % of the peak-to-peak variation of $12 \times 10^8 \text{ kg/s}$. The uncertainties in time varying river discharge and ice melt contribute to a large part of error. In the long term, mass is conserved, and the first term in Equation 5 is negligible. The total ocean surface water flux should balance the total water discharge from continent to ocean. The $\int \nabla \cdot \Theta$ 4 year mean of 10.6 cm/year, computed from outputs of the statistical model, is lower than the climatological value of 12 cm/year given in textbook published 36 years ago (Budyko, 1974) and higher than the climatological river discharge of 8.6 cm/year (Dai and Trenberth, 2002). Large and Yeager (2009) compiled available E and P to give an annual mean of 11 cm/yr. There are, in general, 20 % uncertainties of these hydrologic balances over global ocean (Figure 4).

Based on Green's Theorem, the areal integral of the flux divergence should balance the line integral of flux out of the boundary. The last term of Equation 5 should equal to total water vapor across the coastlines of all continents. Another example of validation by the conservation principle is given by Liu et al. (2006). They first demonstrated the continental water balance in South America (Figure 5). With climatological river discharge ($\int R$) removed from $\int \Theta$ across the entire continental coastline, the residue agrees, both in phase and in magnitude, with monthly rate of mass change ($\int \partial M / \partial t$). The standard deviation of the difference between $\int \partial M / \partial t$ and moisture flux-river discharge is $0.9 \times 10^8 \text{ kg/s}$, which is 7 % of the peak-to-peak variation of $13 \times 10^8 \text{ kg/s}$.



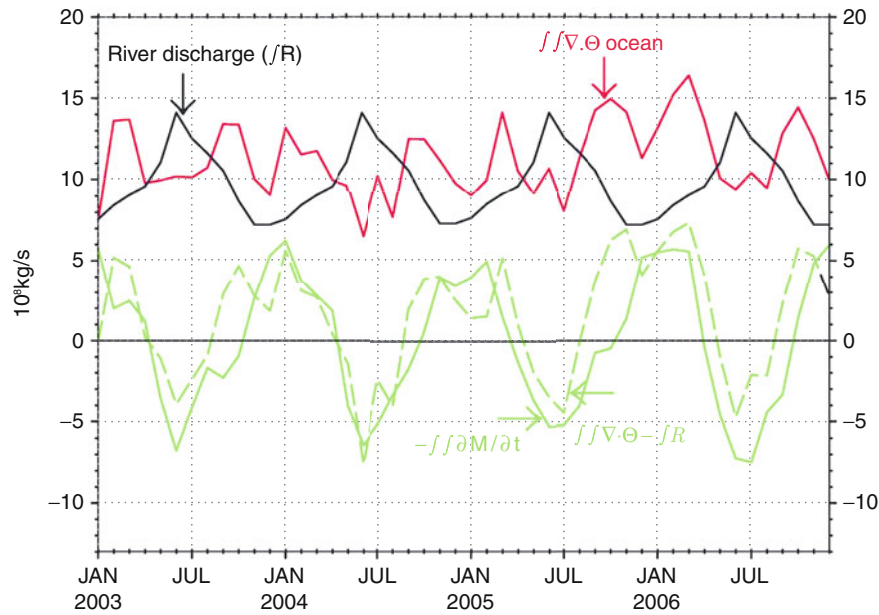
Ocean-Atmosphere Water Flux and Evaporation, Figure 3 Bin-averaged zonal component (a) and meridional component (b) of integrated moisture transport (Θ), derived from satellite data, compared with coincident data computed from rawinsondes.

Marine atmosphere water conservation

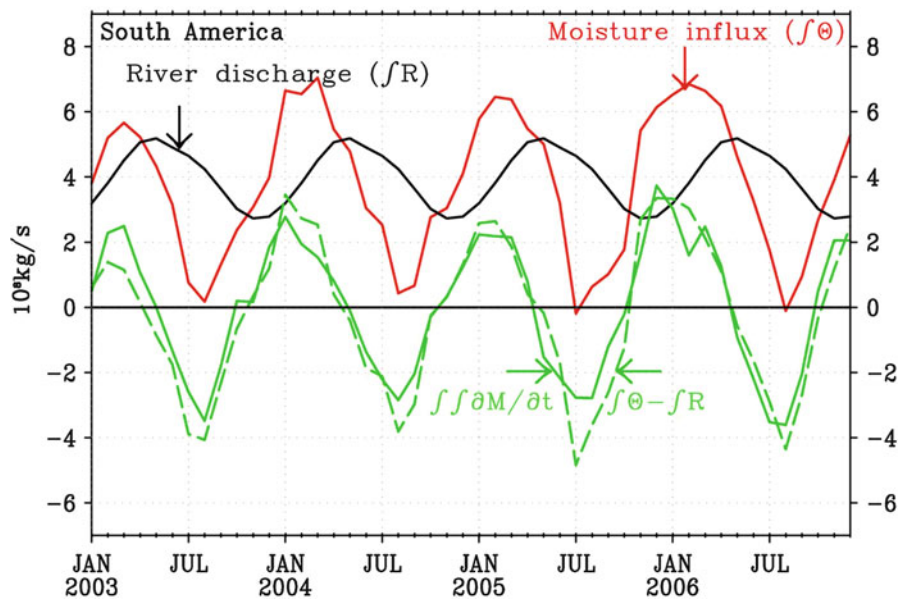
The two flux products should agree with the conservation principle (Equation 1). As an example, the 3 year averages of $\nabla \cdot \Theta$ and E-P are shown in Figure 6. In this example, P is based on TRMM merged data product 3B42, and E is from our direct retrieval from BT. There are general agreements in the magnitude and geographical distribution, but differences in the details. Away from coastal regions, the supply side is larger than the demand side in the tropical southeastern Pacific, tropical south Atlantic, and a region from the Somali coast extending

into the northern Arabian Sea. The demand side is larger than the supply side in the warm pool of the western tropical Pacific and under the Intertropical Convergence Zone (ITCZ). Two operational E products : the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS 3, Andersson et al., 2010) and the Objectively Analyzed air-sea Fluxes (OAFlux) (Yu and Weller, 2007), combined with the same TRMM precipitation are also shown as comparison.

The differences between supply and demand side may reveal regional hydrodynamics. E is the air-sea



Ocean-Atmosphere Water Flux and Evaporation, Figure 4 Annual variation of hydrologic parameters integrated over global oceans.

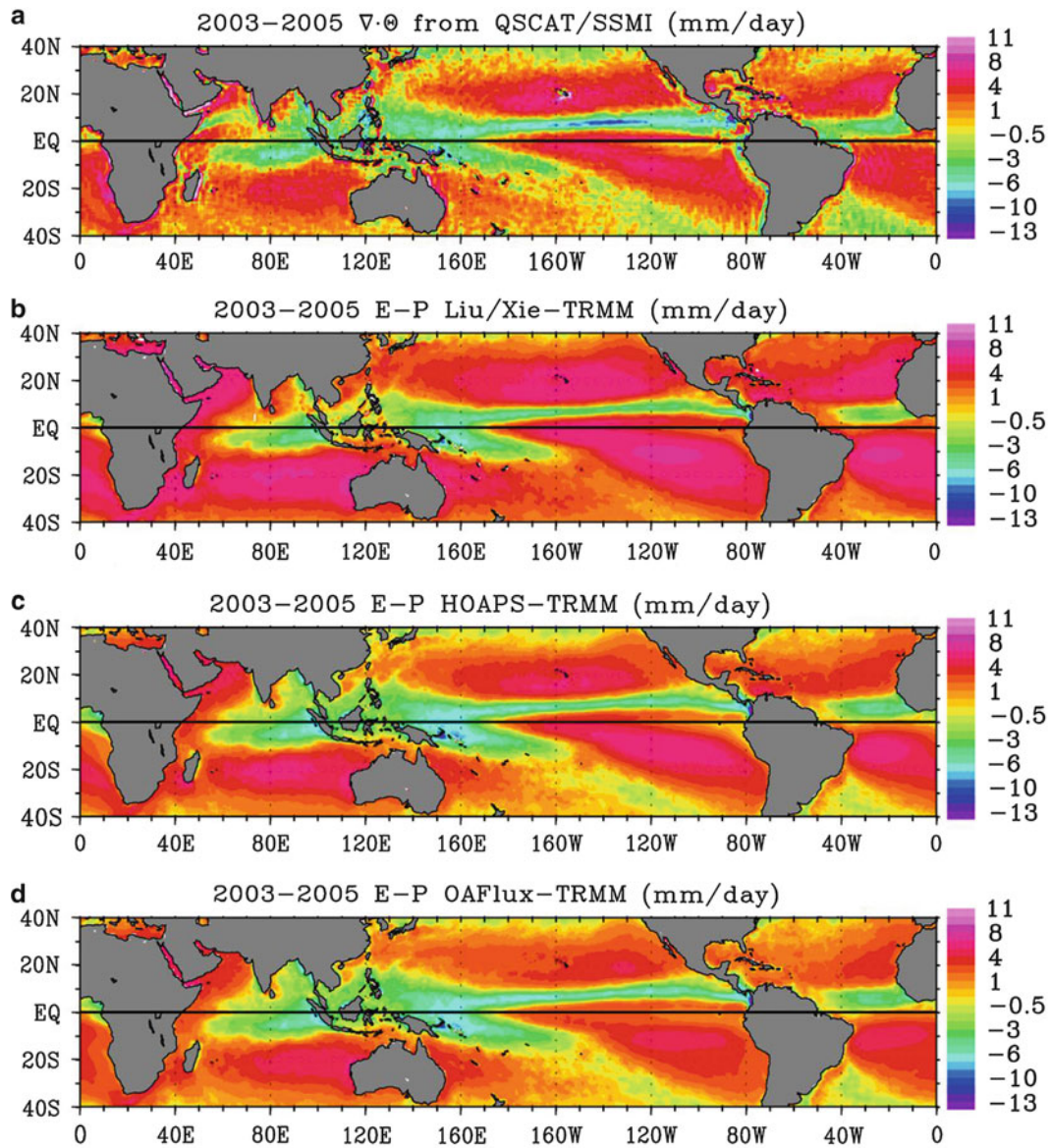


Ocean-Atmosphere Water Flux and Evaporation, Figure 5 Annual variation of hydrologic parameters over South America: mass change rate $\int \partial M / \partial t$ (solid green line), climatological river discharge $\int R$ (solid black line), total moisture transport across coastline into the continent $\int \Theta$ (red line), and $\int \Theta - \int R$ (dashed green line).

exchange of water vapor by turbulence; the small-scale turbulence is largely independent of factors governing large-scale atmospheric circulation (e.g., baroclinicity, Coriolis force, pressure gradient force, cloud entrainment), while Θ is not as sensitive as E to small-scale ocean processes.

Ocean heat and surface salinity conservation

Evaporative cooling is a major variable component of ocean surface heat balance. The LH has been combined with sensible heat flux (SH) and radiative fluxes to provide the net surface thermal forcing of the ocean (Liu and Gautier, 1990; Liu et al., 1994).



Ocean-Atmosphere Water Flux and Evaporation, Figure 6 Three year (2003–2005) annual mean distribution of (a) the divergence of integrated moisture transport, (b) evaporation-precipitation derived from AMSR-E and TRMM, (c) and (d) are the same as (b) except for evaporation from HOAPS 3 and OAF flux.

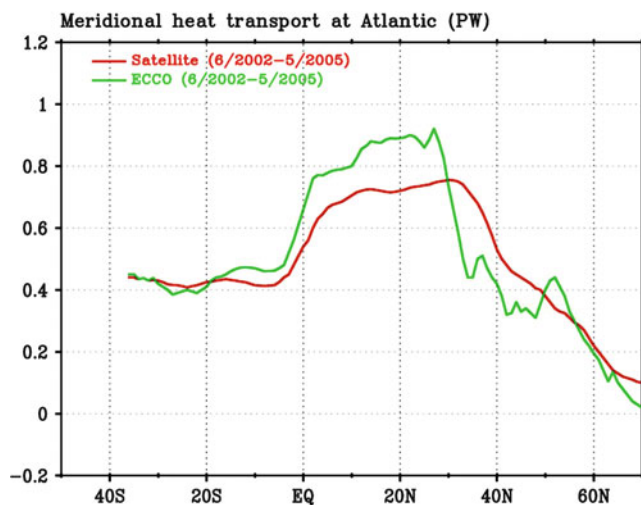
The meridional heat transport (MHT) at a latitude θ is derived by integrating from θ to θ_0 across the width of an ocean basin (x_1 to x_2), the rate of heat content changes subtracting the net surface heat flux,

$$\text{MHT}(\theta) = \int_{\theta}^{\theta_0} \int_{x_1}^{x_2} \left(\frac{\partial H}{\partial t} - \text{SW} + \text{LW} + \text{LH} + \text{SH} \right) dx dy \quad (6)$$

where H is the heat content, SW is the net incoming short-wave radiation flux, and LW is the net outgoing long-wave radiation flux.

The northern end of the ocean basin (θ_0) is treated as enclosed by land. The long term mean meridional heat transport of the major ocean basins have been estimated from the ocean surface fluxes (WCRP, 1982). With recent intense effort to measure the meridional overturning current and the feasibility of measuring H by Argo floats, GRACE, and radar altimeter, we may even examine the temporal variation of the meridional heat transport as a constraint to the LH .

There are large uncertainties in long-term annual mean of MHT compiled in past studies, including those derived from surface flux climatology and from oceanographic



Ocean-Atmosphere Water Flux and Evaporation,
Figure 7 Comparison of annual mean MHT at the Atlantic as a function of latitude. Red curve is calculated using the surface heat balance from satellite observations (SW and LW from the Surface Radiation Budget (SRB), LH and SH from AMSR-E). The green curve is computed from ECCO data.

measurements. Figure 7 shows that the MHT computed from our space-based surface heat flux (red line) is lower than those from the simulation of the Estimating the Circulation and Climate of the Ocean (ECCO) model (green line, Fukumori, 2002) between the equator and 30 °N and higher than ECCO between 30 °N and 50 °N. It agrees with ECCO surprisingly well south of the equator.

The equation of water balance in the upper ocean is

$$\frac{h_0}{S_0} \left(\frac{\partial S}{\partial t} + \mathbf{V} \cdot \mathbf{S} \right) = E - P \quad (7)$$

where V is current and S is salinity in the surface mixed layer with average depth h_0 and average salinity S_0 . Salinity measurements have advanced by the Argo floats and space-based sensors of Aquarius and the Soil Moisture and Ocean Salinity Mission (SMOS). The current velocity has been derived from the displacements of drifters with drogues centered at 15 m depth (Niiler, 2001). Ocean surface currents are also provided by the Ocean Surface Currents Analysis-Real-time (OSCAR) program, using a combination of scatterometer and altimeter data (Lagerloef et al., 1999), at a 5 day and 1° resolution between 70 °S to 70 °N. The ocean surface salinity balance can also be used to put constraints on the accuracy of F . The first term represents the change of storage could be neglected in the long-term mean.

Figure 8 shows the distribution of the surface flux agrees with the distribution of salinity advection in the general features, using Argo data for the 11 year mean and using Aquarius for the 1 year mean.

Summary

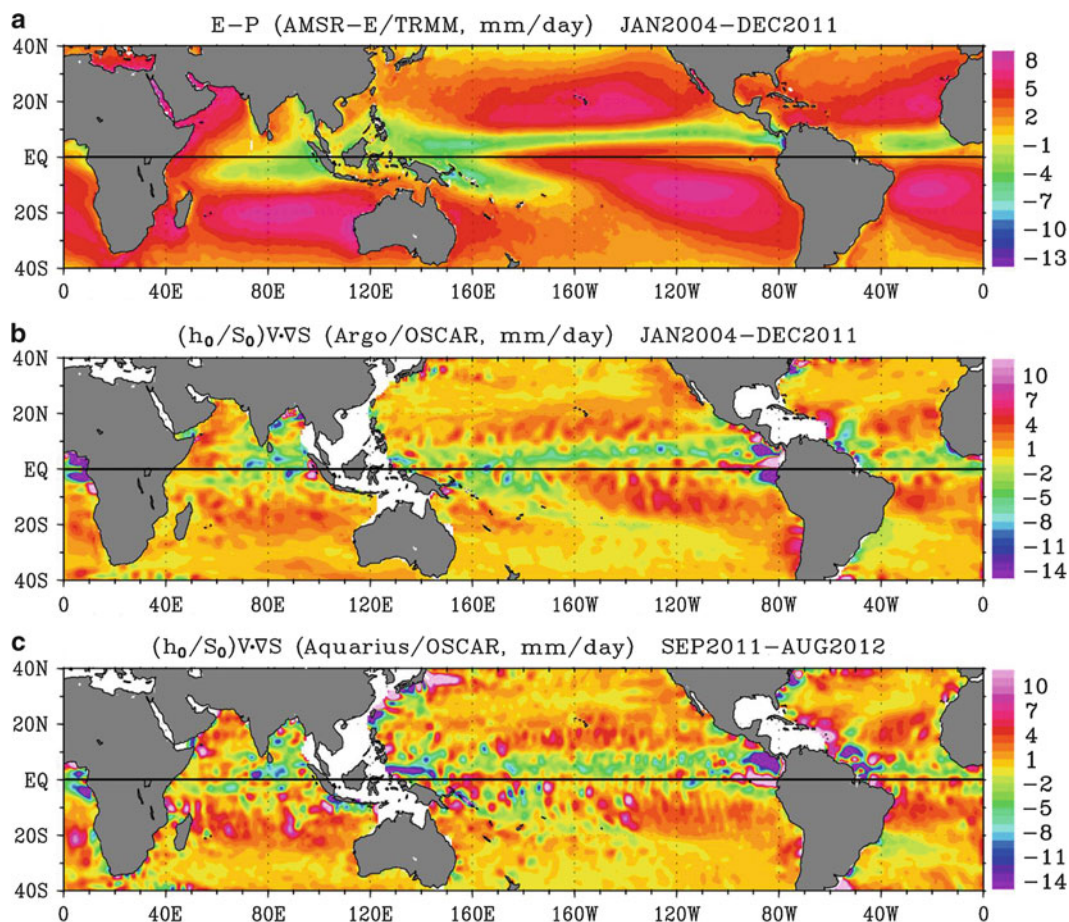
There have been continuous endeavors to estimate E and LH over global oceans using satellite data and based on bulk parameterization of turbulence transport, since Liu and Niiler (1984) successfully estimated the flux by introducing an empirical relation between monthly W and q . With some improvement in this “supply side” approach, a number of data sets have been operationally produced in the past two decades, but large differences among these data sets and between products from satellite data and from reanalysis of operational weather prediction remain (e.g., Curry et al., 2004). We have introduced a new method of direct retrieval of E and LH from the radiances measured by microwave radiometers, which improves the random error of the daily value of LH to 10 % of the dynamic range, as compared with the 19 % error using the methods we pioneered 30 years ago of computing the fluxes from bulk parameters derived from the same radiances.

Evaluations to find the optimal flux product are difficult because of the lack of credible standards (e.g., extensive direct flux measurement). One good constraint to the uncertainties is the closure of the atmospheric water budget, which dictates that $E - P$ should balance $\nabla \cdot \Theta$. The “demand side” approach of estimating Θ and $\nabla \cdot \Theta$ from satellite data serves not only as a credible way to evaluate traditional “supply side” flux products but also to provide the ocean freshwater exchange as a whole, without the need of securing precipitation data separately. The Θ data have been extensively tested in comparison with all available rawinsonde data and products of numerical models. The water flux data, as $\nabla \cdot \Theta$, are also validated through mass conservation using data from GRACE and river discharge climatology; the validation study shows 20 % uncertainties of the seasonal water balance. The feasibility of using upper ocean heat and salinity conservations is also demonstrated with very preliminary results.

There is still much room left for improvement in estimating water flux over global ocean. The new space-based data products, with better spatial and temporal resolution, have many ongoing scientific applications.

Acknowledgment

This report was prepared at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). The Precipitation Measuring Mission, NASA Energy and Water Studies, and the Physical Oceanography Program (through the Ocean Surface Salinity and the Ocean Surface Vector Wind Science Teams) jointly supported this effort. Open data access is provided at <http://airsea.jpl.nasa.gov/seaflux/water-exchange.html>, by the Climate Science Center of JPL as its strategic planning.



Ocean-Atmosphere Water Flux and Evaporation, Figure 8 (a) E-P calculated from AMSR-E and TRMM, (b) salinity advection estimated using Argo and OSCAR data, averaged from 2004 to 2011, and (c) salinity advection from Aquarius and OSCAR data averaged from September 2011 to August 2012.

Bibliography

- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E., 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology*, **4**, 1147–1167.
- Andersson, A., Fennig, K., Klepp, C., Bakan, S., Grassl, H., and Schulz, J., 2010. The Hamburg Ocean atmosphere parameters and fluxes from satellite data - HOAPS-3. *Earth System Science Data*, **2**, 215–234, doi:10.5194/essd-2-215-2010.
- Bourras, D., 2006. Comparison of five satellite-derived latent heat flux products to moored buoy data. *Journal of Climate*, **19**, 6291–6313.
- Bourras, D., Eymard, L., and Liu, W. T., 2002. A neural network to estimate the latent heat flux over oceans from satellite observations. *International Journal of Remote Sensing*, **23**, 2405–2423.
- Brunke, M. A., Zeng, X., and Anderson, S., 2002. Uncertainties in sea surface turbulent flux algorithms and data sets. *Journal of Geophysical Research*, **107**(C10), 3141, doi:10.1029/2001JC000992.
- Budyko, M. I., 1974. *Climate and Life*. New York: Academic.
- Chou, S. H., Atlas, R. M., and Ardizzone, J., 1995. Estimates of surface humidity and latent heat fluxes over oceans from SSM/I data. *Monthly Weather Review*, **123**, 2405–2435.
- Cresswell, S., Ruprecht, E., and Simmer, C., 1991. Latent heat flux over the North Atlantic Ocean—a case study. *Journal of Applied Meteorology*, **30**, 1627–1635.
- Curry, J. A., et al., 2004. Seaflux. *Bulletin of the American Meteorological Society*, **85**, 409–419.
- Dai, A., and Trenberth, K. E., 2002. Estimates of freshwater discharge from continents: latitudinal and seasonal variations. *Journal of Hydrometeorology*, **3**, 660–687.
- Fukumori, I., 2002. A partitioned Kalman filter and smoother. *Monthly Weather Review*, **130**, 1370–1383.
- Hilburn, K. A., 2010. Intercomparison of water vapor transport datasets. *Presented at 17th Conference on Satellite Meteorology and Oceanography and 17th Conference on Air-Sea Interaction*, Annapolis, MD.
- Huffman, G. J., Adler, R. F., Arkin, P. A., Chang, A., Ferraro, R., Gruber, A., Janowiak, J. J., Joyce, R. J., McNab, A., Rudolf, B., Schneider, U., and Xie, P., 1997. The global precipitation climatology project (GPCP) combined precipitation data set. *Bulletin of the American Meteorological Society*, **78**, 5–20.

- Jackson, D. L., Wick, G. A., and Robinson, F. R., 2009. Improved multisensor approach to satellite-retrieved near-surface specific humidity observations. *Journal of Geophysical Research*, **114**, D16303, doi:10.1029/2008JD011341.
- Jones, C., Peterson, P., and Gautier, C., 1999. A new method for deriving ocean surface specific humidity and air temperature: an artificial neural network approach. *Journal of Applied Meteorology*, **38**, 1229–1245.
- Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, **5**, 487–503.
- Kummerow, C., et al., 2000. The status of the tropical rainfall measuring mission (TRMM) after two years in orbit. *Journal of Applied Meteorology*, **39**, 1965–1982.
- Lagerloef, G. S. E., Mitchum, G., Lukas, R., and Niiler, P., 1999. Tropical Pacific near-surface currents estimated from altimeter, wind and drifter data. *Journal of Geophysical Research*, **104**, 23,313–23,326.
- Large, W.G., Yeager, S.G., 2009. The global climatology of an interannually varying air-sea flux data set. *Glim. Dyn*, **33**, 341–364.
- Liu, W. T., 1986. Statistical relation between monthly precipitable water and surface-level humidity over global oceans. *Monthly Weather Review*, **114**, 1591–1602.
- Liu, W. T., 1990. Remote sensing of surface turbulence flux. In Geenaert, G. L., and Plant, W. J. (eds.), *Surface Waves and Fluxes*. Dordrecht: Kluwer, Vol. II, pp. 293–309. Chapter 16.
- Liu, W. T., 1993. Ocean surface evaporation. In Gurney, R. J., Foster, J., and Parkinson, C. (eds.), *Atlas of Satellite Observations Related to Global Change*. Cambridge: Cambridge University Press, pp. 265–278.
- Liu, W. T., and Gautier, C., 1990. Thermal forcing on the tropical Pacific from satellite data. *Journal of Geophysical Research*, **95**, 13209–13217.
- Liu, W. T., and Katsaros, K. B., 2001. Air-sea flux from satellite data. In Siedler, G., Church, J., and Gould, J. (eds.), *Ocean Circulation and Climate*. New York: Academic, pp. 173–1179. Ch. 3.4.
- Liu, W. T., and Niiler, P. P., 1984. Determination of monthly mean humidity in the atmospheric surface layer over oceans from satellite data. *Journal of Physical Oceanography*, **14**, 1451–1457.
- Liu, W. T., and Tang, W., 2005. Estimating moisture transport over ocean using spacebased observations from space. *Journal of Geophysical Research*, **110**, D10101, doi:10.1029/2004JD005300.
- Liu, W. T., Tang, W., and Niiler, P. P., 1991. Humidity profiles over ocean. *Journal of Climate*, **4**, 1023–1034.
- Liu, W. T., Zheng, A., and Bishop, J., 1994. Evaporation and solar irradiance as regulators of the seasonal and interannual variabilities of sea surface temperature. *Journal of Geophysical Research*, **99**, 12623–12637.
- Liu, W. T., Xie, X., Tang, W., and Zlotnicki, V., 2006. Spacebased observations of oceanic influence on the annual variation of South American water balance. *Geophysical Research Letters*, **33**, L08710, doi:10.1029/2006GL025683.
- Miller, D. K., and Katsaros, K. B., 1991. Satellite-derived surface latent heat fluxes in a rapidly intensifying marine cyclone. *Monthly Weather Review*, **120**, 1093–1107.
- Niiler, P., 2001. The world ocean surface circulation. In Siedler, G., Gould, J., and Church, J. (eds.), *Ocean Circulation and Climate-Observing and Modeling the Global Ocean*. London: Academic, pp. 193–204.
- Rienecker, M. M., et al., 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, **24**, 3624–3648.
- Roberts, B., Clayson, C. A., Robertson, F. R., and Jackson, D., 2010. Predicting near-surface atmospheric variables from SSM/I using neural networks with a first guess approach. *Journal of Geophysical Research*, **115**, D19113, doi:10.1029/2009JD013099.
- Santorelli, A., Pinker, R. T., Bentamy, A., Katsaros, K. B., Drennan, W. M., Mestas-Núñez, A. M., and Carton, J. A., 2011. Differences between two estimates of air-sea turbulent heat fluxes over the Atlantic Ocean. *Journal of Geophysical Research*, **116**, C09028, doi:10.1029/2010JC006927.
- Schlüssel, P., Schanz, L., and Englisch, G., 1995. Retrieval of latent heat flux and longwave irradiance at the sea surface from SSM/I and AVHRR measurements. *Advances in Space Research*, **16**, 107–116, doi:10.1016/0273-1177(95)00389-V.
- Schulz, J., Meywerk, J., Ewald, S., and Schlüssel, P., 1997. Evaluation of satellite-derived latent heat fluxes. *Journal of Climate*, **10**, 2782–2795.
- Smith, S. R., Hughes, P. J., and Bourassa, M. A., 2011. A comparison of nine monthly air-sea flux products. *International Journal of Climatology*, **3**, 1002–1027, doi:10.1002/joc.2225.
- Wagner, D., Ruprecht, E., and Simmer, C., 1990. A combination of microwave observations from satellite and an EOF analysis to retrieve vertical humidity profiles over the ocean. *Journal of Applied Meteorology*, **29**, 1142–1157.
- WCRP, 1982. *Report of JSC/CCCO 'Cage' Experiment: A Feasibility Study*. Geneva: World Climate Research Program/World Meteorological Organization.
- WCRP, 1983. *Report of the WMO/CAS Expert Meeting on Atmospheric Boundary, Parameterization Over the Oceans for Long Range Forecasting and Climate Models*. Geneva: World Climate Research Program/World Meteorological Organization.
- Wentz, F.J., and Meissner, T., 2000. AMSR Ocean Algorithm, Version 2, Report number 121599A-1. Remote Sensing Systems, Santa Rosa, CA, pp. 66.
- Xie, X., Liu, W. T., and Tang, B., 2008. Spacebased estimation of moisture transport in marine atmosphere using support vector machine. *Remote Sensing of Environment*, **112**, 1846–1855.
- Yu, L., and Weller, R. A., 2007. Objectively analyzed air-sea heat fluxes (OAFux) for the global ice-free oceans. *Bulletin of the American Meteorological Society*, **88**, 527–539.

OPERATIONAL TRANSITION

Richard Anthes
University Corporation for Atmospheric Research,
Boulder, CO, USA

Synonyms

Research to operations; Technology transfer

Definition

Operational transition is the end-to-end set of processes that lead to the successful development and implementation of a research idea, technology, or observation in an ongoing and useful application (operations). An example is the use of atmospheric observations in daily weather forecasts, which have many applications of benefit to society.

Introduction

Observing Earth through remote sensing and using the observations to increase scientific understanding as well as in various practical applications is critically important for society as it attempts to achieve health, prosperity, safety, and sustainability (NRC, 2007). The use of observations in the research process to increase understanding generally works well, as researchers' careers are dependent upon peer-reviewed publications and other forms of disseminating their research. They are highly motivated to use the observations for discovery and to increase knowledge. However, the process for moving or transitioning the use of observations from research into operations is fraught with peril, and like other forms of technology transfer, the metaphor "crossing the valley of death" is often used (NRC, 2000, 2003). There are various pathways that can result in a successful traverse of the valley, but there are also many challenging hazards and obstacles that can block the pathways so that the transitions fail and opportunities to realize the potential value of research and technologies in operations are lost.

The pathways connecting research and operations are not one-way streets; ideally there is healthy two-way traffic, with operational experience feeding back to the research community, causing additional advances in research and subsequent improved operational processes and products. An example of this productive two-way exchange is the early testing of new atmospheric sounding technologies such as IASI (Infrared Atmospheric Sounding Interferometer) in operational numerical weather prediction models (NRC, 2007). The models, with their use of many other independent observations, often detect errors in new observations or processing algorithms, which provide valuable information to the researchers, who can then diagnose and eliminate the sources of errors.

Most of the research and operational uses of remote sensing of Earth involve NASA on the research side and NOAA and other federal agencies, such as DoD and USGS, on the operational side. Universities and national labs often partner with NASA, NOAA, and other federal agencies on the basic and applied research and to some extent on the transitioning process itself.

There is a growing private sector involvement in providing satellite data, primarily imagery, to paying customers. In many of these enterprises, the private sector carries out the entire transitioning and operational process in order to develop and protect its own intellectual property. In this entry we will focus on government-to-government transitions, since these account for the majority of transitions of remote sensing research to operations.

Transition pathways

Transition pathways are end-to-end sets of processes that lead to a successful transition (NRC, 2003). Pathways generally begin with an idea and involve a basic research phase, a testing for operational utility through a prototype,

and development of a full operational system. Each pathway must include a solid research foundation, often including laboratories, instruments, and other equipment; computers; models; and information technologies. Each step of the pathway must be adequately supported with financial resources; very often excellent ideas are never transitioned into operations because funds are lacking for the transition process and/or continued operations. Above all, people with energy, talent, and vision to see the entire pathway, as well as work together toward the final goal, are needed.

Pathways may be successful through either a "push" of new ideas and opportunities from the research side, a "pull" from the operations side to meet new needs or requirements, or ideally a combination of both. Cultural differences between the research and operational communities may impede a robust transition pathway; researchers may have little interest in applications and operations, and operational communities may have little time or even capability to see and understand the potential value of new technologies. Mechanisms to encourage communication and dialog between the communities involved in all parts of the pathway need to be present, and information should flow in both directions. As described above, feedback from the operational testing and implementation of research data or products to the developers can be very useful in the research process and for improving the technology and products. Test beds may be helpful in the final step before operations, to fully test the technology and associated products, demonstrate their value, and optimize them for routine use in operations.

Transition pathways may occur entirely within the government or private sectors, with or without involvement of the academic sector, or they may involve all three sectors simultaneously. Most aspects of the transition process remain the same in all cases, but the financial support of the transitioning process and the ownership of intellectual property associated with the final operational products may vary considerably.

Bayh-Dole Act of 1980

The Bayh-Dole Act of 1980 was a landmark piece of legislation designed to encourage the transfer of government-funded research done at universities and national laboratories into commercial products for the benefit of society. Prior to the Bayh-Dole Act, most of the new developments and technologies generated with government funding were never patented, but were placed in the public domain, where they languished in the valley of death because industry could not claim any proprietary or competitive advantage. Less than 5 % of patents owned by the government were licensed by industry for development of commercial products. The Bayh-Dole Act permitted ownership of new technologies developed with government funds by the university or lab, so that they could grant exclusive rights to commercial developers. The government retained rights to use the technologies for its own

purposes, but the ability to grant exclusive licenses to private companies provided a valuable incentive to the company to develop the technology for sale and thereby make it available to the public.

The Bayh-Dole Act was successful in encouraging universities to transfer new technologies to the private sector, and the result has been the development of many new products and technologies that benefit the public, as well as the private sector, through creation of jobs and profits. The Bayh-Dole Act also encourages partnerships and collaborations between industry and universities.

Characteristics of successful transitions

Successful transitions from research to operations involve a number of important factors:

1. *Good idea.* The idea must be sound and potential must exist for a useful product. This may seem obvious, but it is fundamental to the entire process. Sometimes attempts are made to transfer weak or incomplete ideas through the research process into operations, only to fail when the idea or its potential fails to be proven in the research or testing process. Many transitions fail because the rigorous research and testing process does not confirm the initial promise of the idea.
2. *Good people.* Talented and dedicated people must be involved at all stages of the transition process. Excellent ideas can take years to implement or not be implemented at all because people fail to understand the potential application of the new research, lack a vision for its use, or are otherwise incapable of carrying the idea through the research and testing phases into operations.
3. *Supporting infrastructure.* Adequate supporting infrastructure, such as laboratory, instruments, computers, and information technology, must be available. As with human resources, this is usually a question of having adequate financial resources.
4. *Financial resources.* Adequate financial resources must be invested into all stages of the transition pathway. This requires investors who have the vision and capability to understand the idea or concept and its potential and the willingness to invest the necessary resources. Lack of adequate resources is probably the number one reason why transitions take too long or fail altogether. This requirement is closely connected with other requirements, notably adequate human resources and supporting infrastructure.
5. *Receptive cultures, people, and responsive organizations.* The culture of the research and operational communities involved in the transition must be supportive and complementary. Organizations must be supportive rather than getting in the way of the transition. Adequate two-way communication and coordination among knowledgeable and dedicated people on the transition pathway must be nurtured and encouraged. Transitions often fail because of cultural differences between the operational and research communities, lack of organizational support, or resistance to change (NRC, 2007). Sometimes new ideas are met with outright hostility because they compete with vested interests and old ways of doing things – the status quo (A quote from Machiavelli is relevant: “There is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things. For the reformer has enemies in all those who profit by the old order, and only lukewarm defenders in all those who would profit by the new order, this lukewarmness arising partly from fear of their adversaries . . . and partly from the incredulity of mankind, who do not truly believe in anything new until they have had actual experience of it.”).
6. *Planning.* Formal planning may not be essential (some ideas are so great that transitions occur in spite of lack of planning), but like any complicated, long journey, good plans – revised as necessary – can make the process more efficient and enable reaching the goal more quickly.
7. *Time is right.* Sometimes great new ideas are proposed, but the scientific knowledge or the technology to support the idea is not ready. In this case, the transition pathway may wait for decades before the process is ready to begin. An excellent example is L.F. Richardson’s concept of numerical weather prediction (Richardson, 1922) which was advanced in the early 1920s but could not be implemented until the advent of computers some 30 years later.

Having laid out all of these characteristics that assist or hinder the transition process, it is worth noting that there are no recipes or formulas that will guarantee success, and serendipity and just pure luck often play a role. Scott Berkun in *Myths of Innovation* puts it this way:

The dirty little secret – the fact often denied – is that unlike the mythical epiphany, real creation is sloppy. Discovery is messy; exploration is dangerous. No one knows what he’s going to get when he is being creative. Creative work cannot fit neatly into plans, budgets, and schedules. Magellan, Lewis and Clark, and Captain Kirk were all sent on missions into the unknown with clear understanding that they might not return with anything, or even return at all.

This quote applies not only to the discovery phase of research to operations pathways, but to the entire transition process.

Examples of successful and unsuccessful transitions

The NRC (2003) report provides a number of examples of successful and unsuccessful transitions of research and technologies to operations. These are generally transitions of research results and technologies from NASA and its partners into operational use by NOAA. Even the successful ones, like the use of infrared sounders from satellites in numerical [weather prediction models](#), took decades to complete. Some, like lightning detection, radar altimetry,

and scatterometry, have yet to be completely transitioned. These examples are instructive in understanding the complex process of research to operations.

Conclusion

In order to achieve maximum societal benefits from observations and research, the research results must be transferred to the operational community so that they can be used on a routine basis for a variety of useful applications. In order to have a successful transition process, it is necessary to have good ideas that are ready to be implemented; talented and dedicated people; adequate supporting infrastructure and financial resources; receptive people, organizations, and cultures; and effective planning. With these pieces of the bridge in place, the valley of death has a very good chance of being successfully crossed.

Bibliography

- Berkun, S., 2007. *Myths of Innovation*. Sebastopol, CA: O'Reilly Media, Incorporated, p. 176.
- NRC, 2000. *From Research to Operations in Weather Satellites and Numerical Weather Prediction-Crossing the Valley of Death*. Washington, DC: National Academy Press, p. 80.
- NRC, 2003. *Satellite Observations of the Earth's Environment – Accelerating the Transition of Research to Operations*. Washington, DC: National Academy Press, p. 163.
- NRC, 2007. *Earth Science and Applications from Space – National Imperatives for the Next Decade and Beyond*. Washington, DC: National Academy Press, p. 428.
- Richardson, L. F., 1922. *Weather Prediction by Numerical Process*. Cambridge, UK: Cambridge University Press.
- U.S. Government Accounting Office GAO, 1998. *Technology Transfer—Administration of the Bayh-Dole Act by Research Universities*. GAO/RCED-98-126. Washington, DC: US General Accounting Office, p. 83.

Cross-references

[Commercial Remote Sensing](#)
[Cost Benefit Assessment](#)
[Emerging Applications](#)
[Emerging Technologies](#)
[Global Programs, Operational Systems](#)
[Mission Operations, Science Applications/Requirements](#)
[Public-Private Partnerships](#)
[Remote Sensing, Historical Perspective](#)
[Weather Prediction](#)

OPTICAL/INFRARED, ATMOSPHERIC ABSORPTION/TRANSMISSION, AND MEDIA SPECTRAL PROPERTIES

Gian Luigi Liberti
 CNR/ISAC, Rome, Italy

Definitions

Optical and infrared. Wavelength range 0.2–20 μm within which main contributions in terms of radiative

processes at the basis of current optical and infrared remote sensing techniques occurs in the Earth atmosphere.

Absorption and transmission are the volume properties that describe the overall effects, of single radiative processes, within a given atmospheric volume, interacting with propagating radiation.

Introduction

Radiative transfer simulations, for example, for remote sensing applications, need to summarize the complexity of radiative processes in a given atmospheric volume due to interaction between propagating radiation and single molecules, aerosols, and hydrometeors, introducing macroscopic or volume properties of the media, such as transmission and absorption (see “[Radiative Transfer Theory](#)”).

Processes responsible for each of the media properties considered are reviewed. The relationships between macroscopic properties and variables describing single components radiative processes are reported. For some case references and links to related codes and/or database will be also given.

Absorption

Absorption is the property of the media that describes the interaction between a propagating electromagnetic (e.m.) wave at a given wavelength and the media characterized by the transformation of the energy associated to the propagating wave into another type of energy (e.g. heat) including e.m. energy at wavelengths different from the original one.

In the atmosphere, due its composition, absorption occurs through two main processes: molecular absorption and absorption by scatterers (aerosols and hydrometeors). In the considered wavelength range, absorption by atomic components can be neglected due both to the reduced energy of photons and to the low concentration of radiatively active pure atomic components in the atmosphere. The most relevant difference in the spectral characteristics of the two absorption processes is due to the fact that molecular absorption is strongly selective (i.e., characterized by large spectral variability), while absorption by condensed matter has smoother spectral features. An example commonly encountered in optical/infrared remote sensing application is the absorption of H_2O that can be found as gas, liquid, or solid in the same optical path. Absorption bands, in addition to the difference in the spectral dependence, are shifted in wavelength; this characteristic allows, for example, the development of remote sensing techniques for cloud particle phase classification in the 1.6 and 10 μm window regions (e.g., Baum et al., 2000).

Absorption due to atmospheric scatterers can be obtained as part of the computation of single-scattering optical properties (see “[Optical/Infrared, Scattering by Aerosols and Hydrometeors](#)”). This section specifically

deals with absorption from molecules and with the relationship between media and single-element (molecule, scatterers) optical properties and the relationship between media optical properties.

Molecular absorption

In the most general case, a molecule in the atmosphere has a total energy E that can be written as the sum of different terms:

$$E = E_{rot} + E_{tr} + E_{vib} + E_{el} \quad (1)$$

where the terms are written in the order of increasing associated energy and refer to the following:

- E_{rot} is the energy associated to rotation of the molecule as a unit body. It depends from the symmetry of the molecules through the number of moments of inertia needed to describe the rotational characteristics of the molecule. Associated energy falls in the far infrared and in the microwave.
- E_{tr} is the translational energy associated to the motion of the molecules. It depends on the temperature of the media (atmosphere) and for typical tropospheric temperatures. For typical temperatures of the Earth atmosphere it falls in the infrared.
- E_{vib} is the energy associated with the vibrational mode. A molecule with N atoms has $3N$ degrees of freedom, of which 3 correspond to the translation. A molecule with N atoms has $3N$ degrees of freedom, of which 3 correspond to the translation of the molecule center of mass and 3 or 2 to the rotation, for a nonlinear or a linear molecule respectively. As a result the remaining degrees of freedom associated to vibrational energy are $3N-6$ (nonlinear molecule) or $3N-5$ (linear molecule).
- E_{el} is the electronic energy. Levels of energy associated to the electronic status are derived from the electronic levels for the atoms forming the molecules decomposed in sublevels due to the influence of the molecular electric field (Stark effect). Associated energy is in the range of the UV and visible.

With the exception the translation energy E_{tr} , the remaining energy terms can take only discrete values that correspond to the transitions between a quantized status of energy to another allowed one. These transitions are regulated by selective rules. Transitions between energy levels associated to a higher form of energy are, in general, associated to transitions between levels of the lower form of energy. A typical example is given by the roto-vibrational absorption bands in the infrared.

Excessive vibrational or electronic excitation results in the dissociation or ionization of the molecule, respectively. Similar unquantized transitions generate continuous absorption spectra that typically occur in the UV region.

For the atmospheric conditions, in addition to the spectra due to transitions between energy levels allowed for a molecule as an isolated entity, collision-induced and

polymer bands should be considered. Collision-induced spectra occur when, for example, a homonuclear diatomic molecule, such as O_2 or N_2 , that has no dipole moment in an isolated state may have, as a consequence of collisions, an induced dipole moment and consequently allowed dipole transitions. Alternatively, depending upon the interaction potential, colliding molecules may form a dimer or larger complex that can survive a few collisions. Such a complex is, from a radiative point of view, a new species with its own vibration-rotation characteristics responsible for the observed polymer spectra. An example is the O_2 - O_2 dimer in the atmosphere.

Line absorption/emission occurs when the energy of a photon ($h\nu$) is equal to the energy required to have an allowed transition from two quantized energetic levels of the molecule, E_1 and E_2 :

$$h\nu = |E_1 - E_2| \quad (2)$$

where h is the Planck constant and ν is the frequency (s^{-1}) of the photon.

According to this definition, absorption (emission) spectra should be formed by strictly monochromatic absorption lines, that is,

$$\sigma_a(\nu) = S\delta(\nu - \nu_0) \quad (3)$$

where σ_a is the absorption cross section, S is the line intensity, and δ is the δ -Kronecker function. Due to the line broadening absorption lines have a finite width, as a consequence:

$$\sigma_a(\nu) = Sf(\nu - \nu_0) \quad (4)$$

where $f(\nu - \nu_0)$ is the generic line shape function that respects the condition:

$$\int_{-\infty}^{+\infty} f(\nu - \nu_0) d\nu = 1 \quad (5)$$

Line broadening occurs due to the following different physical processes:

- *Natural broadening.* Strictly monochromatic absorption is not possible either from the point of view of classical radiation theory due to the damping of any emitted energy or from the quantum theory due to the finite width of quantized energy levels. Resulting line width is, however, from the point of view of current remote sensing applications, negligible being of the order of 10^{-11} cm^{-1} .
- *Collisional broadening.* It depends from the deformation of the molecules, and consequent shift of the energy levels, due to collisions among them. From the kinetic theory of gases it can be derived that the number of collision per unit time is proportional to the ratio $p/T^{1/2}$ where p and T are the pressure and the temperature of the gas volume respectively. Collisional broadening is often referred to as *pressure broadening* because of the direct dependence from pressure.

The characteristics of the collisional broadening depends from the nature of the interacting molecules. It is important to distinguish between: self-broadening due to collisions among molecules of the same species (e.g., $\text{H}_2\text{O}-\text{H}_2\text{O}$) and foreign broadening due to collisions between molecules of different species (e.g., $\text{H}_2\text{O}-\text{N}_2$). This process dominates the line broadening in the troposphere with typical values for line width of the order of 10^{-2} cm^{-1} .

- **Doppler broadening.** It is due to the molecular thermal motion for which a photon emitted by a molecule moving with respect to an observer with a speed V will have a Doppler shift in the frequency proportional to the ratio V/c where c is the speed of light. From Maxwell distribution of velocities, a dependence from the gas temperature $T^{1/2}$ can be derived. In the terrestrial atmosphere at sea level even if temperature can have large values, collisional broadening dominates, and the Doppler half-width is approximately 100 times smaller than the collisional one. At approximately 30–40 km of altitude, depending on the molecule, Doppler broadening equals the collisional one becoming the dominant line broadening process at higher levels.

Line shape, that is, the functional form used to describe the line broadening, is an important issue particularly for the computation of molecular absorption outside the line center wavelength. In particular, the contributions of the far wings sum up in the *continuum*, that is, a sort of very smooth spectral dependence absorption outside the absorption bands. Commonly adopted functional forms are (e.g., Lenoble, 1993) the *Lorentz profile* (for natural and collisional broadening), the *Doppler profile* (for Doppler broadening), and the *Voigt profile* (a combination of the previous two, used to describe the line broadening in for pressure and temperature conditions where Doppler and collisional broadening have comparable values).

In the most general case, remote sensing applications require the computation of the average *absorption (transmission)* $\bar{A}(T)$ within a given spectral interval v around a wavelength value v . For molecular absorption from a single absorbing gas with a density a within the optical path, this can be written as

$$\bar{A}(v) = 1 - \bar{T}(v) = \frac{1}{\Delta v} \int_{v-\Delta v}^{v+\Delta v} [1 - \exp(-a\sigma_a)] dv \quad (6)$$

where if L lines contribute to the absorption in the spectral interval v then

$$\sigma_a(v) = \sum_{j=1}^L \sigma_a^j(v) = \sum_{j=1}^L S^j f^j(v - v_o^j) \quad (7)$$

The above computation can be done using several approaches and corresponding numerical models, depending on the specific application. Relevant

parameters in the selection of the method can be spectral resolution (e.g., radiometry vs. spectrometry) and range (e.g., UV vs. IR), geometry of the observation (limb vs. zenith), objective of the observations (remote sensing of cloud properties, profiling of minor constituents, atmospheric correction for surface remote sensing applications, etc.), and finally logistic considerations (creation of Lookup Table, design of new sensors characteristics, etc.).

Highest spectral resolution computations are obtained with the line-by-line (LBL) approach that is based on considering the contributions, at a given wavelength, from each single absorption line within a certain wavelength range (e.g., the LBLRTM code (Shephard et al., 2009) available at www.rtwweb.aer.com/ or for a review and intercomparison of LBL's models, Tjemkes et al., 2003). Line-by-line computations require detailed information on single-line characteristics (intensity, central wave number, broadening parameters, etc.) that are collected in line atlas. The most widely used and well-documented line atlas is the HITRAN (Rothman et al., 2009), available at www.cfa.harvard.edu/hitran/. The current version of the HITRAN database includes spectral characteristics of 47 molecules including many of their isotopologues. For some molecules or in a certain region of the e.m. spectrum (e.g., UV), absorption properties are given not in terms of single-line properties as derived from spectroscopy but as experimentally derived (i.e., in laboratory inverting the Beer-Bouguer-Lambert law) cross sections.

Computation of molecular absorption through the LBL approach is demanding in terms of computational resources and time. Even considering the increasing availability of computational resources, remote sensing applications that do not require to solve for the single absorption line contributions, as for example radiometry, can apply other methods to compute the effects of molecular absorption. Apart for experimentally based approaches, lower resolution approaches are based still on LBL computations, applied at least once, to compute or to validate the absorption parameters used in the model.

Several solutions have been proposed (see for a review Goody and Young, 1989), for example:

- **Regular models.** These models are based on the assumption that the band is formed by lines spaced at regular intervals of constant (e.g., Elsasser) or variable intensity (e.g., Curtis). With this assumption it is possible to find analytical solutions to the integration over a given wavelength range.
- **Random models.** With this approach lines are assumed to be distributed, in the wavelength range of interest, randomly, with respect to the frequency, following a probability distribution function whose analytical properties allow for a solution to the integral effect of all lines.
- **K-distribution (KD).** The KD method (e.g., Fu and Liou, 1992) is based on the fact that, within a given spectral interval v , radiative variables related to

scattering and thermal emission processes, because of their relatively low spectral variability, can be considered constant. On this basis the integration over frequency [Equation 6](#) is operated changing the variable in distribution of line intensity. Examples of KD-based codes can be found at www.rtwweb.aer.com/rtrtm_frame.html or at <http://snowdog.larc.nasa.gov/cgi-bin/rose/flp200503/flp200503.cgi>.

Finally, the total atmospheric volume absorption, that is, due to the combined effect of molecules and scatterers, is computed, assuming the independence among processes, by summing the absorption cross sections for each species.

Emission

In conditions of *Local Thermodynamic Equilibrium (LTE)*, *Absorption* and *Emission* properties have the same spectral characteristics according to *Kirchhoff's law*. More complex relationships linking absorption to emission properties must be adopted for remote sensing applications strongly based on radiative transfer in the upper stratosphere or higher atmospheric levels (e.g., limb sounding in the upper stratosphere), where *Non Local Thermodynamic Equilibrium (NLTE)* conditions occur (see "[Trace Gases, Stratosphere, and Mesosphere](#)").

Transmission and reflection

For a non-scattering medium (e.g., an atmospheric volume with negligible scattering as a cloud-free layer in the IR) absorption and transmission are related by [Equation 6](#).

In the most general case, in the optical/IR range to fully describe the properties of an atmospheric volume, it is also necessary to introduce the concept of reflection that occurs in the presence of scatterers including molecules. For this reason from the point of view of radiative processes, reflection in the atmosphere depends from elastic scattering, and, in conditions of single scattering (i.e., optically thin layer), it is possible to express the reflection of the layer as a function of single-scattering optical properties.

Absorption deals with the radiative processes that transform the energy of the radiation interacting with the media into other forms of energy including radiation at different wavelengths, for this reason the only relevant geometrical parameter is the optical path of the incoming radiation within the media. Transmission and reflection, dealing with radiation emerging somehow from the volume, need to take into account of the more complex geometry of propagation described by the incident and by emerging radiation. Therefore, the concepts of *direct/specular* and *diffuse* are often introduced to specify particular cases of geometry.

Summary

Radiative processes responsible for the atmospheric volume properties of absorption and transmission have been shortly reviewed, giving examples of methods to

practically account for, in radiative simulation studies. Finally, for some wavelength, in the Optical/IR range and/or remote sensing applications (e.g., optical lighting detection techniques), stimulated emission processes such as *Raman scattering*, *Excited emission*, and *Fluorescence* are the basis for remote sensing techniques.

Bibliography

- Baum, B. A., Soulen, P. F., Strabala, K. I., King, M. D., Ackerman, S. A., Menzel, W. P., and Yang, P., 2000. Remote sensing of cloud properties using MODIS airborne simulator imagery during SUCCESS. II. Cloud thermodynamic phase. *Journal of Geophysical Research*, **105**(D9), 11781–11792.
- Fu, Q., and Liou, K. N., 1992. On the correlated k-distribution method for radiative transfer in non-homogeneous atmospheres. *Journal of the Atmospheric Sciences*, **49**(22), 2139–2156.
- Goody, R. M., and Young, Y. L., 1989. *Atmospheric Radiation. Theoretical Basis*, 2nd edn. New York: Oxford University Press.
- Lenoble, J., 1993. *Atmospheric Radiative Transfer*. Hampton, VA: A. Deepak.
- Rothman, L. S., et al., 2009. The HITRAN 2008 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **110**, 533–572.
- Shephard, M. W., Clough, S. A., Payne, V. H., Smith, W. L., Kireev, S., and Cady-Pereira, K. E., 2009. Performance of the line-by-line radiative transfer Model (LBLRTM) for temperature and species retrievals: IASI case studies from JAIVEx. *Atmospheric Chemistry and Physics Discussions*, **9**, 9313–9366.
- Tjemkes, S., et al., 2003. The ISSWG line-by-line intercomparison experiment. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **77**, 433–453.

Cross-references

[Earth Radiation Budget, Top-of-Atmosphere Radiation](#)
[Lightning](#)
[Limb Sounding, Atmospheric](#)
[Optical/Infrared, Radiative Transfer](#)
[Optical/Infrared, Scattering by Aerosols and Hydrometeors](#)
[Radiation, Electromagnetic](#)
[Radiative Transfer, Theory](#)
[Reflected Solar Radiation Sensors, Multiangle Imaging](#)
[Thermal Radiation Sensors \(Emitted\)](#)
[Trace Gases, Stratosphere, and Mesosphere](#)
[Trace Gases, Troposphere - Detection from Space](#)
[Ultraviolet Sensors](#)
[Water Vapor](#)

OPTICAL/INFRARED, RADIATIVE TRANSFER

Knut Stamnes
 Stevens Institute of Technology, Castle Point on Hudson,
 Hoboken, NJ, USA

Definition

Radiation. Electromagnetic energy in the solar and terrestrial part of the spectrum.

Short-wave radiation. Electromagnetic energy in the solar part of the spectrum.

Long-wave radiation. Electromagnetic energy in the terrestrial part of the spectrum.

Radiative transfer. Transport of electromagnetic energy in a scattering and absorbing medium.

Blackbody. A cavity, filled with matter of any kind, that absorbs all radiation incident upon it and emits radiation in accordance with Planck's spectral distribution law.

Introduction

Remote sensing relying on optical/infrared radiation depends on the transport of radiation through a scattering and absorbing medium. For example, remote sensing of the Earth from space requires interpretation of the electromagnetic signal either backscattered or emitted from the atmosphere and the underlying surface. Successful retrieval of surface properties is hampered by the presence of atmospheric molecules and aerosol particles that contribute significantly to the measured signal. To retrieve atmospheric and surface parameters from space, realistic simulations of the radiative transfer process in the coupled atmosphere-surface system are required.

Sources of atmospheric radiation: atmospheric windows

The bulk of the Earth's atmosphere (99 % by mass) consists of molecular nitrogen and oxygen, which are radiatively inactive (homonuclear, diatomic) molecules that have negligible impact on absorption and emission, but are responsible for molecular (Rayleigh) scattering or radiation. Trace amounts of polyatomic molecules are responsible for atmospheric absorption and emission of radiation in several hundred thousands of individual spectral lines arising from rotational and vibrational transitions (Goody and Yung, 1989; Liou, 2002; Thomas and Stammes, 1999). The atmospheric contribution to the top-of-the-atmosphere radiance depends on scattering and absorption by molecules and aerosol particles. As shown in Figure 1, there is little overlap between the radiation spectra of the Sun and the Earth. Therefore, except for applications where the region of overlap (3–4 μm) is of special interest, we may treat the two spectra separately.

The gaseous absorption is strongly wavelength dependent, whereas the aerosol effect varies smoothly and relatively slowly with wavelength. Figure 1 demonstrates that there are certain "atmospheric windows" in which gaseous absorption is very small. Thus, radiances measured in these window regions are suitable for retrieval of aerosol and surface properties, and satellite instruments are usually designed to take advantage of them.

Radiative processes and the radiative transfer equation

The spectral radiance I_ν is the energy d^4E per unit area, per unit solid angle, per unit frequency, and per unit time:

$$I_\nu = \frac{d^4E}{\cos\theta dA dt d\omega dv} [\text{W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \text{Hz}^{-1}]$$

where θ is the angle between the surface normal $\hat{n}$ and the direction of propagation $\hat{\Omega}$.

Absorption, scattering, and extinction by molecules and particles

A beam of light incident on a thin atmospheric layer will interact with matter, and if the layer has thickness ds , then the light is attenuated so that the differential loss in radiance is $dI_\nu = -kI_\nu ds$, where k is called the extinction coefficient. Integration along the beam path through the layer yields

$$I_\nu(s, \hat{\Omega}) = I_\nu(0, \hat{\Omega}) \exp[-\tau_\nu(s)]. \quad (1)$$

Here $\hat{\Omega}$ denotes the propagation direction of the beam, and the dimensionless extinction optical path or opacity along the path s is given by $\tau_\nu(s) \equiv \int_0^s ds' k(\nu)$. Attenuation of a light beam in a specific direction can be caused by either absorption or scattering.

Angular scattering by molecules and particles

The scattered power per unit area per steradian in a particular direction of observation divided by the power per unit area of the incident beam is called the angular scattering cross section, $\sigma(\Theta)$. The angle Θ between the directions of incidence $\hat{\Omega}'$ and observation $\hat{\Omega}$ is given by $\cos\Theta = \hat{\Omega}' \cdot \hat{\Omega} = \cos\theta' \cos\theta + \sin\theta' \sin\theta \cos(\phi' - \phi)$. Thus, in spherical coordinates θ' and ϕ' are the polar and azimuthal angles of the incident beam and θ and ϕ those of the scattered beam. We define the phase function as the normalized angular scattering cross section:

$$p(\cos\Theta) = \frac{\sigma(\cos\Theta)}{\int_{4\pi} d\omega \sigma(\cos\Theta)}$$

which has the normalization $\int_{4\pi} d\omega p(\cos\Theta)/4\pi = 1$.

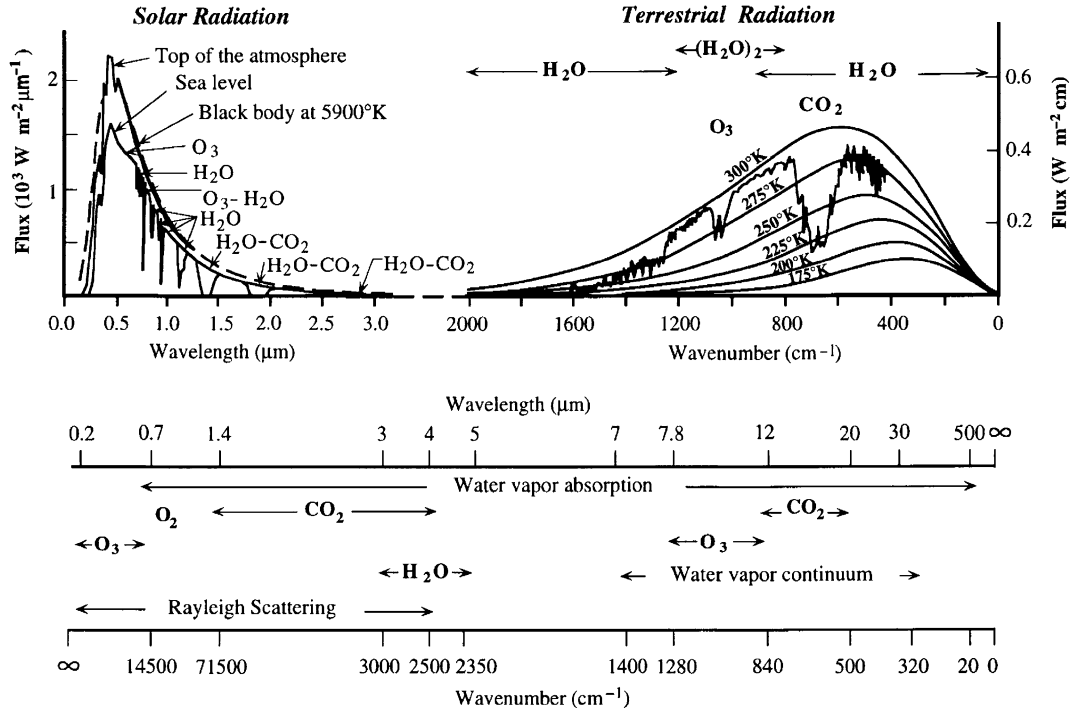
Long-wave absorption and emission by the surface and the atmosphere

The Earth's atmosphere is in contact with land and ocean surfaces. Their strong continuous absorption in the IR allows them to be treated as blackbodies, which emit radiation according to Planck's law:

$$I_\nu^{BB} = B_\nu(T) \equiv \frac{m_r^2}{c^2} \frac{2h\nu^3}{(e^{h\nu/k_B T} - 1)}$$

where h is Planck's constant, c is the speed of light, m_r is the real index of refraction, and k_B is Boltzmann's constant.

The spectral *directional emittance* is defined as the ratio of the energy emitted by a surface of temperature T_s to the energy emitted by a blackbody at the same frequency and temperature $\in(\nu, \hat{\Omega}, T_s) \equiv I_{\nu_e}^+(\hat{\Omega}) \cos\theta / B_\nu(T_s) \cos\theta d\omega = I_{\nu_e}^+(\hat{\Omega}) / B_\nu(T_s)$. A surface for which $\in$ is unity for all $\hat{\Omega}$ and ν , is a *blackbody*, by definition. Similarly, we define the *spectral directional absorptance* as the ratio of



Optical/Infrared, Radiative Transfer, Figure 1 Spectral distribution of solar (shortwave) and terrestrial (longwave) radiation fields. Also shown are the approximate shapes and positions of the scattering and absorption features of the Earth's atmosphere.

absorbed energy to incident energy of the beam $\alpha(v, -\hat{\Omega}', T_s) \equiv I_{v\alpha}^-(\hat{\Omega}') \cos \theta' d\omega' / I_v^-(\hat{\Omega}') \cos \theta' d\omega' = I_{v\alpha}^-(\hat{\Omega}') / I_v^-(\hat{\Omega}')$. Kirchhoff's law states that for an opaque surface $\alpha(v, -\hat{\Omega}, T_s) = \epsilon(v, \hat{\Omega}, T_s)$, whereas for an extended medium the thermal volume emission coefficient j_v^{th} is

$$j_v^{th} = \alpha(v) B_v(T).$$

Short-wave surface reflection and transmission

For an angular beam of radiation with radiance $I_v^-(\hat{\Omega}')$ within a cone of solid angle $d\omega'$ around $\hat{\Omega}'$, the energy incident on a flat surface with normal $\hat{n}$ is $I_v^-(\hat{\Omega}') \cos \theta' d\omega'$, where θ' is the angle between $\hat{\Omega}'$ and $\hat{n}$. Denoting by $dI_{vr}^+(\hat{\Omega})$ the radiance of reflected light leaving the surface within a cone of solid angle $d\omega$ around the direction $\hat{\Omega}$, we define the *bidirectional reflectance distribution function*, or BRDF, as the ratio of the reflected radiance to the energy in the incident beam $\rho(v, -\hat{\Omega}', \hat{\Omega}) \equiv dI_{vr}^+(\hat{\Omega}) / I_v^-(\hat{\Omega}') \cos \theta' d\omega'$. Adding the contributions to the reflected radiance in the direction $\hat{\Omega}$ from beams incident on the surface in all downward directions, we obtain the total reflected radiance

$$I_{vr}^+(\hat{\Omega}) = \int dI_{vr}^+(\hat{\Omega}) = \int_{2\pi} d\omega' \cos \theta' \rho(v, -\hat{\Omega}', \hat{\Omega}) I_v^-(\hat{\Omega}').$$

The BRDF, or in short the *reflectance*, plays a central role in the remote sensing of planetary surfaces and is important for the correct assessment of their albedo.

Purely diffuse reflection occurs at microscopically irregular surfaces, while purely specular reflection occurs when the surface is perfectly smooth, like a mirror. If the reflected radiance from a surface is completely uniform, i.e., independent of the angle of observation, it is called a *Lambert surface*. The BRDF for a Lambert surface is independent of both the direction of incidence and the direction of observation. Then the reflectance simplifies to $\rho(v, -\hat{\Omega}, \hat{\Omega}) = \rho_L(v)$, where ρ_L is the *Lambert reflectance*. Specular reflection from and transmission through a smooth dielectric surface can be calculated from *Snell's law* and *Fresnel's equations*, given the optical constants of air and the dielectric material.

The equation of radiative transfer

For an atmospheric layer of thickness ds that not only attenuates but also emits radiation, we find that the differential change in radiance becomes $dI_v = -kI_v ds + j_v ds$ or $dI_v/d\tau_s = -I_v + S_v$ where the source function, S_v , is the sum of two terms: $S_v = j_v^{th}/k(v) + j_v^{sc}/k(v) = S_v^{sc} + S_v^{th}$. Here the volume emission coefficient for scattering is $j_v^{sc} = \sigma(v) \int_{4\pi} (d\omega/4\pi) p(\hat{\Omega}', \hat{\Omega}) I_v(\hat{\Omega}')$, while (from Kirchhoff's law) that for thermal emission is just $j_v^{th} = \alpha(v) B_v(T)$. Thus, the complete radiative transfer

equation, which includes both multiple scattering and absorption, becomes (Thomas and Stamnes, 1999; Zdunkowski et al., 2007)

$$\frac{dI_v}{d\tau_s} = -I_v + [1 - a(v)]B_v(T) + \frac{a(v)}{4\pi} \int_{4\pi} d\omega' p(\hat{\Omega}', \hat{\Omega}) I_v(\hat{\Omega}') \quad (2)$$

where $a(v) = \sigma(v)/k(v)$, the *single-scattering albedo*. Equation 2 has a simple physical interpretation: the term on the left-hand side is the change in the radiance along the path ds , the first term on the right-hand side is the loss of radiation due to extinction, the second term is the gain due to emission, and the third term is the gain due to multiple scattering.

Conclusion

Solutions of the radiative transfer equation allow for computation of the radiance that would be measured by a remote sensing instrument. They also allow for computation of weighting functions required in classic inverse methods designed to retrieve atmospheric and surface properties from remote measurements (Rodgers, 2000).

Bibliography

- Goody, R. M., and Yung, Y. L., 1989. *Atmospheric Radiation, Theoretical Basis*, 2nd edn. New York: Oxford University Press.
- Liou, K.-N., 2002. *An Introduction to Atmospheric Radiation*. New York: Academic.
- Rodgers, C., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Singapore: World Scientific.
- Thomas, G. E., and Stamnes, K., 1999. *Radiative Transfer in the Atmosphere and Ocean*. Cambridge: Cambridge University Press.
- Zdunkowski, W., Trautmann, T., and Bott, A., 2007. *Radiation in the Atmosphere: A Course in Theoretical Meteorology*. Cambridge: Cambridge University Press.

Cross-references

[Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties](#)
[Optical/Infrared, Scattering by Aerosols and Hydrometeors](#)

OPTICAL/INFRARED, SCATTERING BY AEROSOLS AND HYDROMETEORS

Gian Luigi Liberti
 CNR/ISAC, Rome, Italy

Definitions

Optical/infrared range of wavelengths $0.2 \div 20 \mu\text{m}$ where the range somehow takes into account of the probability, referred to the Earth atmosphere, to have scattering of natural radiation (i.e., solar or from Earth-atmosphere-ocean thermal emission) from particulate suspended in the atmosphere.

Elastic *scattering* of electromagnetic (e.m.) radiation occurs when the propagating e.m. wave traverses an inhomogeneity in terms of propagation properties. Only elastic scattering is considered in this entry.

Hydrometeors are used for particulate solid and liquid formed principally by water, while *aerosols* include all remaining (i.e., nonwater) liquid or solid particles suspended in the atmosphere.

Introduction

Elastic scattering of electromagnetic (e.m.) radiation occurs when the propagating e.m. wave traverses an inhomogeneity (scatterer) in terms of e.m. propagation properties. The effect of the scatterer depends on the length (relative to the wavelength in the propagating medium) of the portion of path, within the discontinuity, of the propagating e.m. wave and on the propagating properties of the scattering medium. In the atmosphere the propagating medium is the vacuum, therefore, atmospheric particulate acts by delaying the propagation of the e.m. wave.

Within the optical/infrared ($0.2 \div 20 \mu\text{m}$) range of wavelengths, molecules are small compared to the wavelengths and therefore their elastic scattering effect either is negligible or can be described by the Rayleigh scattering (Rayleigh, 1871). Because of the relative stability of atmospheric composition and the weak dependence from the nature of the scattering molecules (e.g., Snee and Ubachs, 2005), scattering from molecules is often assumed as known or dependent only from the density or pressure. This allows to use it as background/calibration signal in some remote sensing application. Examples are Mie-Rayleigh lidar calibration (Fiocco and Grams, 1964), aerosols extinction profiling (Ansmann et al., 1990), cloud detection (Buriez et al., 1997), and detection of absorbing aerosols in the UV (Herman et al., 1997). Quantitative information, through remote sensing techniques, can be retrieved for liquid and solid particulates in the atmosphere for which, for some wavelength in the range of interest, the effect of scattering depends on the physical and chemical characteristics of the scattering particulate. This condensed matter is generally divided into two major classes according with the dominant composition:

- *Hydrometeors* that include condensed water vapor particulate both in liquid as well as solid form. Hydrometeors represent the basic components of clouds and precipitation.
- *Aerosols* that include the mixture of micron-sized liquid or solid particles, except water, suspended in the air.

Because of the processes of cloud droplet nucleation and of aerosol hygroscopic growth, there is obviously an overlap in the two classes of atmospheric particulate. A typical example is the transition from haze, considered as a type of aerosols, to fog that is a type of cloud. In addition to the dominant particle composition, another relevant

aspect in terms of scattering properties is that hydrometeors are in general larger than aerosol particles.

From a historical point of view, most of the optical phenomena that occur in the atmosphere such as rainbows, halos, glories, and coronas are generated by scattering by hydrometeors, and their observation and interpretation have been documented since the antiquity (e.g., [Aristotle ~350 BC](#)).

Physical parameters from which scattering properties depend

In order to be able to describe the set of basic single-scattering optical properties (SSOPs, e.g., extinction coefficient, single-scattering albedo, and scattering matrix) at a given wavelength for a single-particulate element, the following properties need to be known:

Size/shape (or habits) and orientation to allow the computation of the physical path, within the particulate of the e.m. radiation. Two main classes are defined to describe these physical properties of the particulate: *spherical* (for which a single parameter, i.e., the radius a , is sufficient to describe the whole physical aspect of the particulate) and *nonspherical*. Distinction between optical properties of spherical and nonspherical constitutes the basis, in some cases, of remote sensing techniques to identify the nature of the scattering particulate, as for example, the analyses of lidar backscatter ratio (for a review, see Sassen, [2005](#)) or POLDER cloud phase algorithm (Goloub et al., [2000](#)). Liquid particulate is generally considered as spherical. This assumption holds up to the size of precipitating droplets up to a radius of about 140 μm above which the deformation becomes large, and droplets are better approximated by oblate or oblate with a flat bottom spheroids.

Particularly important is the description of particle habit for ice hydrometeors. In fact, although the basis of solid water is hexagonal, the overall shape depends on a combination of factors (temperature, available humidity, rate, and process of growth). This means that ice crystal habit contains useful information on the characterization of the observed cloud volume but, on the other hand, it makes it difficult to develop remote sensing inversion algorithms, for example, for the estimation of optical thickness and cloud microphysical properties (e.g., Baum et al., [2005](#)).

A minimum set of variables to quantitatively describe size and shape of particles is based on the following parameters:

- *Size parameter*: $x = 2\pi a/\lambda$ where λ is the wavelength of e.m. radiation in the surrounding medium and a is, for spherical particles, the radius while for nonspherical can be another characteristic particle size (e.g., semi major dimension, radius of surface, or volume equivalent sphere).
- *Aspect ratio*: ratio of the maximum to the minimum particle dimensions.

With reference to orientation (obviously relevant only for nonspherical particles), most of quantitative applications assume randomly oriented particles. This is the case when atmospheric turbulence dominates the internal motion of particles in a given atmospheric volume. In some case, as, for example, in the upper troposphere, intense laminar air fluxes can be responsible for orientation of particles. Similarly, orientation of hydrometeors (i.e., plates, snowflakes) in absence of strong atmospheric motion can be determined by the aerodynamic properties of the particles (e.g., Breon and Dubrulle, [2004](#)). The optical phenomenon of sun dogs, for example, is due to orientation of ice particles.

Chemical composition from which the propagation properties of the electromagnetic wave depends

This is represented at a given wavelength by a value of the complex refractive index where the real part is responsible for the delay of the e.m. wave, while the imaginary part is responsible for the absorption of energy from the e.m. wave. The complex refractive index also shows dependence from the temperature that can be already relevant for the range of possible atmospheric temperatures, for example, when focusing on particular cases of particulate as, for example, supercooled water droplets. Taking correctly into account of temperature dependence may be even more important for remote sensing applications for the study of planetary atmospheres. If the refractive index is different for different states of polarization, the composite medium is birefringent and/or rotates the plane of polarization. Examples of birefringent media that can be found in atmospheric particulate are ice (Takano and Liou, [1989](#)), quartz (SiO_2), and hematite.

An up-to-date compilation of tabulated complex refractive indices and relevant documentation for basic components of atmospheric particulate can be found in the HITRAN (Rothman et al., [2009](#)) package (www.cfa.harvard.edu/hitran/). As an example, the 2008 version (V13.0) includes refractive indices of water, supercooled water, ice, sodium chloride (NaCl), ammonium sulfate $(\text{NH}_4)_2\text{SO}_4$, sea salt, water-soluble aerosol, carbonaceous aerosol, volcanic dust, meteoric dust, dust-like aerosol, aqueous sulfuric and nitric acid, solid hydrates (i.e., nitric acid mono-, di-, and tri-drates), organic nonvolatile aerosol, crustal material (e.g., quartz, hematite, and sand), and ternary $\text{H}_2\text{SO}_4/\text{HNO}_3/\text{H}_2\text{O}$ droplets at low temperatures as components of polar stratospheric clouds (PSCs).

Single-scattering optical properties (SSOPs) of single particles

SSOPs for single-scattering particles can be computed with *numerical methods* based on the exact or approximate solution of Maxwell's equations of the wave propagating within the discontinuity or can be measured experimentally.

Several numerical methods to compute the SSOP have been developed. They can be classified in terms of:

- Size (or more correctly, size parameter) of the particle
- Shape of the particle, for example, spherical, regular solid, combination of regular solids, axially symmetric, and irregular shape
- Difference between the real part of the refractive index of the medium and of the scatterer (often referred as soft and hard particles)
- Whether it is an approximate or an exact solution

Basic textbooks dedicated to methods to compute SSOP for a large set of cases are Van de Hulst (1957), Kerker (1969), Bohren and Huffman (1983), Mishchenko et al. (2002), and Quinten (2011). A specific review of numerical and experimental methods for computation of SSOP for nonspherical particles is in the book edited by Mishchenko et al. (2002).

Programs to perform computation of SSOP have been developed and distributed even before the existence of the Web (e.g., (Dave, 1968), MIEV0 (Wiscombe, 1980), BHMIE and BHCYL (Bohren and Huffman, 1983)). Libraries containing source programs, executable programs, or online computations are now widely diffused. The SCATTERLIB Web site (atol.ucsd.edu/scatlib/scatterlib.htm) by Piotr Flatau contains links to a large set of codes for SSOP computation. In some cases, radiative transfer model packages (e.g., LIBRADTRAN www.libradtran.org, Mayer and Kylling, 2005) include libraries of programs for the computation of SSOP.

The selection of the method/package used to compute SSOP, apart for the applicability to the class of particles of interest, can be driven by the output parameters generated by the computation and in particular the possibility to compute the whole set of elements of the scattering matrix.

Measurement of SSOP is an alternative method to characterize optical properties of aerosols and hydrometeors or can be used to validate results obtained by numerical models. Because of the complexity of experimental setup to perform such type of measurements, relevant literature is by far less rich than literature on theoretical and numerical methods. In the optical/infrared range, two possible approaches have been used:

- *Direct measurement.* The experimental setup includes a light source, polarization devices for the incident and for the scattered radiation to select the scattering matrix element, a detector of the scattered light, and a scattering volume where the scattering particulate is illuminated. Due to their reduced dimension of scatterers of interest in this wavelength range, it is difficult to accurately characterize the size and shape of scattering particles. Another strong limitation of this approach is that the arrangement of light source and detector precludes scattering angle measurements close to 0° and 180° scattering angle with the consequence that the estimation of the scattering coefficient, obtained by

integrating over the whole range of scattering angle, needs extrapolation or integration with numerically obtained results. Similarly, the extinction coefficient in case of small particle is difficult to be measured with finite aperture detectors because of the presence of forward-scattered light. A review on direct measurements can be found in Hovenier (2000).

- *Analog microwave (MW).* Owing to the fact that scattering properties, for a given refractive index, depend uniquely by the size parameter, difficulties in characterizing the scattering particles can be reduced by using larger wavelengths (i.e., MW) for which the scattering object becomes a few order of magnitudes larger than for direct measurement, allowing for easier technical solutions to control size, shape, and orientation. Another advantage of the analog MW approach is that the range of allowed scattering angle can include the exact forward direction. A limitation of this technique is that measurements can be done only for one particle size, shape, and orientation at the time with consequent large effort to obtain more commonly required results for polydispersions or other kind of ensemble-averaged results. A review on MW analog to light-scattering measurements can be found in Gustafson (2000).

SSOP of polydispersion, mixtures, and atmospheric volumes

Processes responsible for the presence in the atmosphere of aerosols and hydrometeors generate particles with different size (*polydispersion*), even in the case of homogeneous chemical composition. The concept of *size distribution* (SD) is introduced to describe the population of a polydispersion of atmospheric particulate as a function of the size. SDs of atmospheric scatterers are generally expressed with a single or with a combination of analytic functions (e.g., Deepak and Box, 1982). More commonly used are lognormal (Davies, 1974), power law (Junge, 1963), Gamma and Modified Gamma (Deirmendjian, 1964, 1969), and exponential (Marshall and Palmer, 1948). The SD for a given atmospheric scatterer depends on the process generating it; for this reason differences among the functions are mostly due to the type of scatterer: aerosols, ice/water nonprecipitating hydrometeors, and ice/water precipitating hydrometeors. In addition, the numbers of parameters needed to fully describe the SD (typically ranging from 2 to 4) is another important factor in the selection of the SD function. In general, the analytical function is asymmetric with respect to the average being bounded inferiorly by allowing only positive values of the particle size, while most of atmospheric processes of interest allows for a small probability of having large particles. SD may be expressed in terms of radius, surface, or volume according to the application. For example, for computation of mass transport, the volume representation gives a more direct information.

In principle, to describe the optical properties of atmospheric particulate in a given air volume, it is necessary

to know size, shape, orientation, and chemical composition of each single particle. With a similar approach, the knowledge of optical properties is clearly a problem with a huge number of unknowns. In order to be able to practically take into account the direct radiative effects of aerosols (a similar approach, apart for the chemical composition, is done for clouds), basic *aerosol components* have been identified, on the basis of experimental observations that are directly linked with the source processes responsible for generation of particulate in the atmosphere. A generic atmospheric volume is then assumed to contain a mixture of single aerosol components. On the basis of observations, *aerosol models* are defined as typical combinations of aerosol components that represent both the presence of sources and the effects of atmospheric processes (i.e., transport, aging) on the particulate generated by the sources.

A first well-documented and widely diffused review and summary of previous results on aerosols and fog optical properties in terms of models is due to Shettle and Fenn (1979). The developed database derived from this review activity has been the basis for aerosol database input in the set radiative transfer models (RTM) developed by the Air Force Geophysics Laboratory (LOWTRAN, MODTRAN, FASCODE) as well as from other RTMs.

Successively, a reference publication for aerosol models has been the book from D'Almeida et al. (1991) that in addition, to the Shettle and Fenn (1979) models, contained in an updated review of aerosol components and models. Moreover, one of the objectives of the volume was to define a global climatology of aerosol scattering properties for a set of given wavelengths. The information contained in D'Almeida et al. (1991) has been used to produce database and packages for computation of SSOP for mixture of components and/or models, for example, OPAC (Hess et al., 1998a) or Levoni et al. (1997).

Aerosol components are often classified on the basis of the interaction with atmospheric water vapor in:

- Hygroscopic particles (e.g., NaCl, nitrates, sulfates) which act as nuclei for the condensation of water vapor responding to changes in atmospheric relative humidity (RH) by absorbing and releasing water vapor as RH changes
- Nonhygroscopic particles (e.g., dust, carbon soot, hematite) which are relatively passive components responding to a wet ambient by accumulating a film of water at the surface

Such classification is important for the computation of optical properties because it determines the kind of mixing. *External mixing* implies that each component of a given aerosol class is represented by a different substance with its own size distribution and complex refractive index. The resultant SSOPs of the mixture are obtained as appropriate weighted averages of the SSOP of each component. *Internal mixing* implies that a single particle is a result of mixture of components, for example,

in a solution. In this case, the refractive index of the resulting mixture and the size of the obtained particle must be computed to compute the SSOP. Other kinds of mixing are, for example, *coated particles* (e.g., water droplet condensating over a soot particle) or touching or aggregated particles, that is, where the particle resulting from the mixture is formed by single particles with their own composition but physically in contact.

Similar to what is available for aerosols, database of optical properties for clouds has been also developed and distributed (e.g., COPE, Hess et al., 1998b).

Summary

Elastic scattering by aerosols and hydrometeors in the optical/infrared range of wavelengths is at the basis of a wide set of remote sensing techniques, both active and passive as well as ground based or from flying platforms. Further improvement of our knowledge of the radiative effects of scatterers in the atmosphere requires a better characterization of the angular distribution of scattering from nonspherical particles, the absorption properties of atmospheric scatterers in the wide range of wavelengths of interest for Earth radiation budget. This improves the capability of remote sensing techniques, particularly in terms of profiling and estimating size distribution parameters at a global scale. It is reasonable to expect, in the near future, an increase of occurrence of active remote sensing instrumentation on satellite missions. Nevertheless, because of the nature of the scattering process, the angular- and polarization-derived information measurable from passive remote sensing instruments has not yet been fully exploited.

Bibliography

- Ansmann, A., Riebesell, M., and Weitkamp, C., 1990. Measurement of atmospheric aerosol extinction profiles with a Raman lidar. *Optics Letters*, **15**, 746–748.
- Aristotle, 350 B.C. *Meteorology*. III.2 (Trans. Webster, E. W.). Available from World Wide Web: www.ebooks.adelaide.edu.au/a/aristotle/meteorology/
- Baum, B. A., Yang, P., Heymsfield, A. J., Platnick, S., King, M. D., Hu, Y. X., and Bedka, S. T., 2005. Bulk scattering properties for the remote sensing of ice clouds. 2: narrowband models. *Journal of Applied Meteorology*, **44**, 1896–1911.
- Bohren, C. F., and Huffman, D. R., 1983. *Absorption and Scattering of Light by Small Particles*. New York: Wiley.
- Breon, F. M., and Dubrulle, B., 2004. Horizontally oriented plates in clouds. *Journal of the Atmospheric Sciences*, **61**(23), 2888–2898.
- Buriez, J.-C., Vanbauce, C., Parol, F., Goloub, P., Herman, M., Bonnel, B., Fouquart, Y., Couvert, P., and Seze, G., 1997. Cloud detection and derivation of cloud properties from POLDER. *International Journal of Remote Sensing*, **18**, 2785–2813.
- D'Almeida, G. A., Koepke, P., and Shettle, E. P., 1991. *Atmospheric Aerosols, Global Climatology and Radiative Characteristics*. Hampton: A Deepak.
- Dave, J. V., 1968. Subroutines for computing the parameters of the electromagnetic radiation scattered by a sphere. Report No. 320–3237. Palo Alto: IBM Scientific Center. IBM Order No. 360D-17.4.002.

- Davies, C. N., 1974. Size distribution of atmospheric particles. *Journal of Aerosol Science*, **5**(3), 293–300.
- Deepak, A., and Box, J. P., 1982. Representation of aerosol size distribution data by analytic models. In Deepak, A. (ed.), *Atmospheric Aerosols*. Hampton: Spectrum Press, pp. 79–109.
- Deirmendjian, D., 1964. Scattering and polarization properties of water clouds and hazes in the visible and infrared. *Applied Optics*, **3**, 187–196.
- Deirmendjian, D., 1969. *Electromagnetic Scattering on Spherical Polydispersions*. New York: Elsevier.
- Fiocco, G., and Grams, G., 1964. Observations of the Aerosol Layer at 20 km by Optical Radar. *Journal of the Atmospheric Sciences*, **21**, 323–324.
- Goloub, P., Herman, M., Chepfer, H., Riedi, J., Brogniez, G., Couvert, P., and Seze, G., 2000. Cloud thermodynamical phase classification from the POLDER spaceborne instrument. *Journal of Geophysical Research*, **105**, 14747–14759.
- Gustafson, B. A. S., 2000. Microwave analog to light-scattering measurements. In Mishchenko, M. I., Hovenier, J. W., and Travis, L. D. (eds.), *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications*. San Diego, CA: Academic, pp. 355–365.
- Hansen, J. E., and Travis, L. D., 1974. Light scattering in planetary atmospheres. *Space Science Reviews (Springer Netherlands)*, **16**(4), 527–610.
- Herman, J. R., Barthia, P. K., Torres, O., Hsu, N. C., Seftor, C. J., and Celarier, E., 1997. Global distribution of UV-absorbing aerosols from Nimbus 7/TOMS data. *Journal of Geophysical Research*, **102**, 16911–16922.
- Hess, M., Koepke, P., and Schult, I., 1998a. Optical properties of aerosols and clouds: the software package OPAC. *Bulletin of the American Meteorological Society*, **79**, 831–844.
- Hess, M., Koelmeyer, R. B. A., and Stammes, P., 1998b. Scattering matrices of imperfect hexagonal ice crystals. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **60**, 301–308.
- Hovenier, J. W., 2000. Measuring scattering matrices of small particles at optical wavelengths. In Mishchenko, M. I., Hovenier, J. W., and Travis, L. D. (eds.), *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications*. San Diego, CA: Academic, pp. 355–365.
- Junge, C. E., 1963. *Air Chemistry and Radioactivity*. New York: Academic Press.
- Kerker, M., 1969. *The Scattering of Light and Other Electromagnetic Radiation*. New York: Academic.
- Levoni, C., Cervino, M., Guzzi, R., and Torricella, F., 1997. Atmospheric aerosol optical properties: a data base of radiative characteristics for different components and classes. *Applied Optics*, **36**(30), 8031–8041.
- Marshall, J. S., and Palmer, W. M. K., 1948. The distribution of raindrops with size. *Journal of Meteorology*, **5**, 165–166.
- Mayer, B., and Kylling, A., 2005. Technical note: the libRadtran software package for radiative transfer calculations – description and examples of use. *Atmospheric Chemistry and Physics*, **5**, 1855–1877.
- Mishchenko, M. I., Hovenier, J. W., and Travis, L. D. (eds.), 2000. *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications*. San Diego, CA: Academic.
- Mishchenko, M. I., Travis, L. D., and Lacis, A. A., 2002. *Scattering, Absorption, and Emission of Light by Small Particles*. Cambridge: Cambridge University Press. Available online from World Wide Web: www.giss.nasa.gov/staff/mmishchenko/books.html.
- Quinten, M., 2011. *Optical Properties of Nanoparticle Systems: Mie and beyond*. John Wiley & Sons.
- Rayleigh, L. (J. W. Strutt), 1871. On the scattering of light by small particles. *Philosophical Magazine* **41**, 447–454.
- Rothman, L. S., et al., 2009. The HITRAN 2008 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **110**, 533–572.
- Sassen, K., 2005. Polarization lidar. In Weitkamp, C. (ed.), *Lidar*. New York: Springer, pp. 19–42.
- Shettle, E. P., and Fenn, R. W., 1979. Models for the aerosol lower atmosphere and the effects of humidity variations on their optical properties, Tech. Rep. Tr-79-0214. Hanscom, MA: Air Force Geophysics Laboratory.
- Sneep, M., and Ubachs, W., 2005. Direct measurement of the Rayleigh scattering cross section in various gases. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **92**, 293–310.
- Takano, Y., and Liou, K. N., 1989. Solar radiative transfer in cirrus clouds. Part I: single-scattering and optical properties of hexagonal ice crystals. *Journal of the Atmospheric Sciences*, **46**(1), 3–19.
- Van de Hulst, H. C., 1957. *Light Scattering by Small Particles*. New York: Wiley (also Dover, New York).
- Wiscombe, W. J., 1980. Improved Mie scattering algorithms. *Applied Optics*, **19**, 1505–1509.

Cross-references

[Aerosols](#)
[Air Pollution](#)
[Calibration, Optical/Infrared Passive Sensors](#)
[Cloud Liquid Water](#)
[Cloud Properties](#)
[Land-Atmosphere Interactions, Evapotranspiration](#)
[Lidar Systems](#)
[Ocean-Atmosphere Water Flux and Evaporation](#)
[Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties](#)
[Radiation, Electromagnetic](#)
[Radiation, Multiple Scattering](#)
[Radiation, Polarization, and Coherence](#)
[Radiation, Volume Scattering](#)
[Reflected Solar Radiation Sensors, Multiangle Imaging](#)
[Reflected Solar Radiation Sensors, Polarimetric](#)
[Water and Energy Cycles](#)

P

PATTERN RECOGNITION AND CLASSIFICATION

Björn Waske¹ and Jón Atli Benediktsson²

¹Institute of Geodesy and Geoinformation,
University of Bonn, Bonn, Germany

²Faculty of Electrical and Computer Engineering,
University of Iceland, Reykjavik, Iceland

Definition

The classification of remote sensing images and the corresponding generation of land cover maps are perhaps the most common applications in remote sensing. In general, the aim of a land cover classification is the assignment of each pixel within the imagery to a specific information class (e.g., forest areas). In general, this is performed by methods of machine learning and pattern recognition. Pattern recognition can be defined as a technique to classify data (patterns) based either on a priori knowledge or statistical information extracted from the patterns.

Introduction

During the last decades, remote sensing became a valuable and important tool to monitor the Earth. Overall it had a significant impact on the acquisition and analysis of environmental data, and the manner how the planet is observed was revolutionized (Rosenqvist et al., 2003). Nowadays remote sensing imagery and corresponding products, such as land cover maps, are helping to support environmental monitoring systems and decision making in the context of different multilateral environmental treaties, such as the Kyoto Protocol and the European initiative Global Monitoring for Environment and Security,

GMES (Peter, 2004; Rosenqvist et al., 2003; Backhaus and Beule, 2005).

The classification of remote sensing data and, thus, the generation of land cover maps are the most common applications. Consequently, the development of adequate classification methods is an ongoing research topic. This entry briefly reviews well-known image classification methods as well as other state-of-the-art methods. However, a detailed discussion of all available algorithms and concepts, which are introduced in the field of pattern recognition and remote sensing, would be beyond the scope of this entry. The interested reader is referred to one of the several textbooks, e.g., Richards and Jia (2006) and Duda et al. (2002). An overview to the classification of multisource remote sensing data is given in Chapter x (Waske and Benediktsson 2014).

Classification of remote sensing data

General definitions

The data, the remote sensing imagery, consist of pixels (i.e., picture elements) and provide different characteristics at different wavelength (in case of multispectral and hyperspectral imagery). When dealing with land cover classifications, we aim at the differentiation between several land cover classes, and a classification algorithm is used to differentiate the types of patterns.

A pattern can be assumed as a unique structure or attribute (e.g., the pixel values), which is capable to describe a specific phenomenon (e.g., a land cover type). During the classification, unknown patterns (i.e., pixel) are assigned to a predefined class or they are combined to unknown clusters, depending on their similarity. An n -dimensional pattern (or pixel) in the context of remote sensing is represented by an n -dimensional random

variable or feature vector X , where n denotes the number of available features (e.g., image bands):

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (1)$$

where x_i , $i = 1, \dots, n$ corresponds to the i -th measurement (the measurement from the i -th wavelength band or sensor channel) on a given ground resolution element. The classification is usually done in a so-called *feature space* (Richards and Jia, 2006).

The field of pattern recognition and classification is highly linked with the area of machine learning. Machine learning refers to the development of methods that optimize their performance iteratively by learning from our data. When performing a land cover classification, we are usually dealing with descriptive models, which enable to distinguish between different classes of patterns. Various classifier concepts have been introduced, which can be sorted into different groups, such as supervised, unsupervised, parametric, and nonparametric methods (Waske and Benediktsson 2014).

Most approaches belong to the group of supervised classification, where a set of labeled samples is available. The objective is to predict y , the class membership of a new and unknown sample x , using the mapping function f . Machine learning theory aims at estimating the functional f from some prior data as the labeled training data.

In contrast to this, in *unsupervised* classification, only the data are available, without any class membership information. In this case, the classification algorithm describes how the data are clustered within the feature space and is enabled to identify unknown structures, such as natural groups within the feature space. Consequently, pixels that belong to different clusters are more dissimilar to each other compared to pixels within the same clusters (Jain et al., 2000). In addition to unsupervised and supervised approaches, *semi-supervised* techniques have been introduced (Shahshahani and Landgrebe, 1994; Jackson and Landgrebe, 2002). In this case, the learning is based not only on the training samples but also on some additional data without any ground truth knowledge. With that, the classification accuracy has been showed to increase, compared to a supervised classification. Besides the dichotomy of supervised and unsupervised, classifiers can also be grouped into *parametric* and *nonparametric* techniques. The widely used supervised maximum likelihood classifier is a parametric approach. This approach is based on the assumption that the probability density function for each class is multivariate, and often a Gaussian distribution is assumed. Contrary to this approach, nonparametric methods are not constrained to any assumptions on the distribution of input data. Hence, techniques such as neural networks, decision trees, and

support vector machines can be applied, even if the class-conditional densities are not known or cannot be estimated reliably. Thus, in the context of multitemporal and multisource data sets or high-dimensional imagery with a limited number of training samples, such techniques can be more adequate in terms of the classification accuracy.

Unsupervised techniques

In general, unsupervised techniques or cluster algorithms assign pixels to specific classes without having a prior knowledge of the classes, by measuring similarities between different pixels in a learning process which most often is iterative. After measuring the similarity between pixels, the algorithm merges the individual pixels to groups. Although different, similar measurements have been introduced in the literature, the Euclidean distance is perhaps the widely used measure when remote sensing data are clustered.

It is worth to underline that the clustering of a particular data can result in different potential cluster combinations, e.g., if the algorithm is initialized with different values. Thus, some kind of evaluation is required to assess the quality of the generated clusters, such as the squared error measure (Richards and Jia, 2006). The cumulative distance of each pixel to its cluster center is determined individually within each cluster and finally the total sum over all clusters is calculated.

In contrast to supervised methods, unsupervised approaches have the advantage that they enable the image classification without having prior knowledge (i.e., training data). On the other hand, the algorithm does not provide any final membership decisions, and in general, the analyst must assign a specific class label to each cluster. Thus, the quality of the final classification might be affected by the knowledge and experience of the analyst.

Well-known cluster methods include the k-means and the ISODATA (e.g., Duda et al., 2001; Richards and Jia, 2006). The latter optimizes the individual clusters during the learning process by splitting and merging operations. Both approaches assign each pixel to only one group, whereas concepts such as fuzzy c-mean clustering provide a degree of membership to all clusters for each pixel.

Supervised methods

As mentioned above, a pixel (or pattern) in the context of remote sensing is represented by an n -dimensional vector X . This representation in the feature space $\mathcal{R}^n$ is used by a classification algorithm to assign unknown pixels to one of c land cover classes $\Omega = \{\omega_i\}$, for $i = 1, \dots, c$, using training data. The differentiation between individual classes is based either on the similarity to a certain class or by decision boundaries, which are constructed in the feature space.

Maximum likelihood classification

In the statistical framework, a two-class classification problem can be formulated as

$$P(\omega_i|x) \underset{\omega_j}{\overset{\omega_i}{>}} P(\omega_j|x), \quad (2)$$

with $P(\omega_i|x)$ being the *posteriori* (or conditional) *probability* that the true class of x is ω_i . Following Equation 2, the sample x is classified to class ω_i if $P(\omega_i|x) > P(\omega_j|x)$. Equation 2 provides the basic concept for the well-known maximum likelihood classifier (MLC).

However, this requires the knowledge of $P(\omega_i|x)$, which is generally unknown and needs to be estimated from the training data. Following the well-known Bayes theorem, the posteriori probabilities can be described by prior and the class-conditional probabilities and Equation 2 can be rewritten as follows:

$$p(\omega_i)p(x|\omega_i) \underset{\omega_j}{\overset{\omega_i}{>}} p(\omega_j)p(x|\omega_j), \quad (3)$$

with $p(x|\omega_i)$ as the *class-conditional* probability density function and $p(\omega_i)$ the prior probabilities. The prior probability describes the probability that class ω_i is present in our image. Let us assume that it is known that 40 % of the image to be classified is covered by class ω_i . Then, the prior would be $p(\omega_i) = 0.4$, indicating that in this case, we would have a 40 % probability of finding a pixel x belonging to class $p(\omega_i)$.

The approach assumes that our different feature vectors have specific probability densities $p(x|\omega_i)$, which are conditioned on the observed land cover classes. Therefore, a pixel x belonging to class ω_i is assumed to be an observation, which is taken randomly from the class-conditional probability density function $p(x|\omega_i)$. Consequently, differences between two classes ω_i and ω_j result in differences between the corresponding likelihoods (density functions) $p(x|\omega_i)$ and $p(x|\omega_j)$ (Duda et al., 2001).

To determine the likelihoods from the training data, it is usually assumed that the distributions follow the form of a multivariate normal (Gaussian) model. It is important to note that other models can be assumed, but many processes can be described by a normal model. Moreover, this type of model simplifies the approach, because the model only has two parameters and is described by the mean and the covariance matrix of the training samples:

$$p_i(x|\omega_i) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_i|}} \times \exp \left[-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) \right] \quad (4)$$

with μ_i as the mean vector from class ω_i , covariance matrix Σ_i , and the inverse Σ_i^{-1} , and T as transpose

operator. By taking logarithms of both sides in Equation 3 and following Equation 4, the (Gaussian) maximum likelihood classifier can be finally described by

$$(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i) + \ln |\Sigma_i| \underset{\omega_j}{\overset{\omega_i}{>}} (x - \mu_j)^T \Sigma_j^{-1} \times (x - \mu_j) + \ln |\Sigma_j|, \quad (5)$$

under the assumption that all classes have equal prior probabilities.

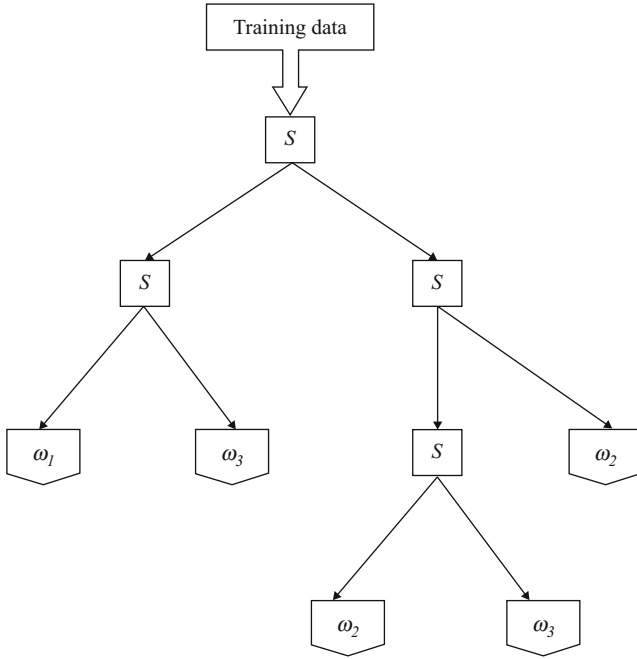
Decision tree classifiers

Decision trees (DT) are a nonparametric method, which can handle different data sets, including categorical variables. The relative structural simplicity of DT and the relatively short training time (compared to computationally complex approaches) are some advantages of a DT classifier (Friedl and Brodley, 1997; Pal and Mather, 2003). In addition, the DT classification scheme allows a direct interpretation of class membership decisions with respect to the impact of individual features. Although a standard DT may be considered limited under some circumstances, the general concept is interesting and the classifier performance in terms of accuracies can be further increased by classifier ensembles or multiple classifier systems (Briem et al., 2002; Gislason et al., 2006; Waske and Braun, 2009).

A general introduction on DT is, e.g., given by Safavian and Landgrebe (1991) and a brief overview is given below. During the construction of the DT (i.e., the classifier training), the training set is successively separated into an increasing number of smaller, more homogenous groups. A DT consists of a root node, which includes all samples, internal nodes with a split rule, and the final leaf nodes, which refer to the different classes. This hierarchical concept is different from other classification approaches, which usually use the entire features space at once and make a single membership decision per class. Moreover, the training process may lead to a splitting of thematic classes (i.e., land cover classes) into several final nodes (Figure 1).

Most applications are based on binary DT, which use, contrary to a multivariate DT, only one feature for the construction of each split node. To determine the split rule, the classifier aims at generating the descendent nodes as pure as possible. A node is assumed to be pure, when all pixels within that node belong to the same class, resulting in an (im)purity value of 0. In contrast, the maximum (im)purity value is reached, if all samples are equally distributed over the classes. A common measurement criterion in this context is Breiman et al. (1984), *Gini index*, which is described as follows (Duda et al., 2001):

$$Gini(S) = \sum_{i=1}^c p_{oi}(1 - p_{oi}), \quad (6)$$



Pattern Recognition and Classification, Figure 1 Schematic flowchart of a decision tree, for a three-class example. At each internal node, there is a split rule S to successively separate the data into smaller groups. Finally, the data ends up in one of the leaf nodes, which refer to a specific land cover class.

with c as the number of classes and $p_{\omega i}$ as the probability of class ω_i at node S :

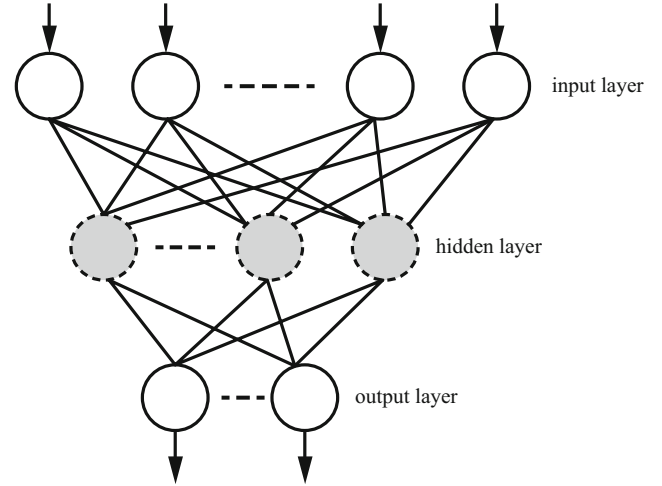
$$p_{\omega i} = \frac{l_{\omega i}}{L}, \quad (7)$$

where $l_{\omega i}$ the number of samples belonging to class ω_i and L as the total number of samples within the training set. The Gini index is used to determine the largest homogeneous group within the training data and discriminate it from the rest. The impurity of all potential groups is summed, and the split rule that causes the maximum reduction in impurity, i.e., which generates the most homogenous node, is selected.

To avoid complex tree structures and thus over-fit the training data, different approaches are considered when dealing with DT. Pre-pruning methods stop the tree induction, e.g., when the number of samples within a node becomes very small or the improvement in the accuracy is too low. Post-pruning methods, on the other hand, are applied after a DT is fully grown. They eliminate inefficient and weak branches and consequently generate a more compact classifier (Esposito et al., 1997).

Neural networks

Neural networks have been used for diverse remote sensing applications, including the classification of



Pattern Recognition and Classification, Figure 2 Schematic diagram of a neural network, with one hidden layer.

multisource data sets and high-dimensional imagery (Benediktsson et al., 1990, 1995; Paola and Schowengerdt, 1995; Bruzzone et al., 2004). As other machine learning approaches, e.g., support vector machines and decision trees, neural networks have the advantage that they are not based on any underlying statistical model of the data, the Gaussian maximum likelihood classifier, for example.

A neural network can be considered as an interconnection of neurons, where each neuron receives input signals, representing the activity at the input or the momentary frequency of neural impulses delivered by another neuron to this input (Bishop, 1995). The function at a neuron o_i can be described as:

$$o_i = Cf(w_i x - \vartheta_i), \quad (8)$$

where C is a constant, x a vector of inputs, and f a nonlinear function, the threshold function, which takes the value 1 for positive argument and 0 for negative arguments. The w_i are weights and ϑ_i are thresholds, with $i = 1, \dots, c$, where c is the number of (land cover) classes.

An adaptive training process is used to determine the weights, using a set of training samples as input. For each sample, the network gives an output response, which is compared to the desired and known output. This information is used to modify the weights in the neural network. When the accuracy cannot be increased any further or the error is reduced to a predefined threshold, the adaptive process ends. A neural network with one hidden layer can be used to differentiate linearly separable data, whereas two or more layers enable the definition of more complex decision boundaries. A schematic overview of a fully connected feedforward neural network is shown in Figure 2.

The backpropagation algorithm (or the multilayer perceptron network) is the best-known neural network algorithm in the context of classification of remote sensing data (for a review see Paola and Schowengerdt, 1995). This multilayer network enables the discrimination between classes, which are not linearly separable. Nevertheless, the approach has some serious drawbacks, such as among others the slow convergence of the algorithm. In contrast to this, radial basis function (RBF) networks can overcome some of the problems with the backpropagation and have a much faster training time (Bishop, 1995; Bruzzone and Prieto, 1999).

Support vector machines

Support vector machines (SVM) have been used for successfully for the classification of diverse remote sensing data sets (Huang et al., 2002; Foody and Mathur, 2004; Melgani and Bruzzone, 2004). The nonparametric approach is particularly interesting for classifying multisource and high-dimensional data sets (Pal and Mather, 2006; Waske and Benediktsson, 2007).

The SVM differentiates two classes by fitting an optimal separating hyperplane to the training samples of two classes in a multidimensional feature space. It can be assumed as an approximate implementation of Vapnik's structural risk minimization principle (Vapnik, 1999). In different remote sensing studies, the SVM has performed efficiently with small training sets, even when classifying high-dimensional imagery (Pal and Mather, 2006). Nevertheless, a sufficient number of training samples are required to ensure the availability of adequate samples for the classifier training (Foody and Mathur, 2004). Moreover, the classifier performance of the SVM can also be affected by the curse of dimensionality in high-dimensional feature spaces and with small or suboptimal training sets (Waske et al., 2010).

The SVM is able to separate complex (e.g., multimodal) class distributions in high-dimensional feature spaces by mapping the data with nonlinear kernel functions. Thus, for linearly not-separable cases, the input data are mapped into a high-dimensional space, wherein the newly distributed samples enable the fitting of a linear hyperplane. Whereas the SVM was originally developed as a supervised classifier, the concept was extended by semi-supervised concepts (Chi and Bruzzone, 2007; Ghoggali et al., 2009) and used for classifier ensembles (Waske et al., 2010).

A detailed introduction on the general concept of the SVM is given in Burges (1998); an overview in the context of remote sensing can be found, e.g., in Huang et al. (2002), Melgani and Bruzzone (2004), and Plaza et al. (2009).

For a binary classification problem in a n -dimensional feature space $\mathcal{R}^n$, the training vectors are given by $x_i \in \mathcal{R}^n, i = 1, 2, \dots, L$, with L training samples and a vector $y \in \mathcal{R}^L$ with the corresponding class labels

$y_i \in \{+1, -1\}$. The SVM aims to solve the following primal problem:

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^L \xi_i \\ \text{subject to} \quad & y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, \dots, L. \end{aligned} \quad (9)$$

Its dual is

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, L \\ & y^T \alpha = 0, \end{aligned} \quad (10)$$

with e as the vector of all ones, C as regularization parameter, Q is an L by L positive semidefinite matrix, and $Q_{ij} = y_i y_j K(x_i, x_j)$ and $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ are the kernel. Thus, the training vectors x_i are implicitly mapped into a higher-dimensional feature space by ϕ . The class membership is based on the sign of the final decision function $f(x)$:

$$f(x) = \sum_{i=1}^L y_i \alpha_i K(x_i, x_j) + b. \quad (11)$$

The kernel strategy enables to work within the newly transformed feature space, without explicitly knowing ϕ , but only the kernel function. A widely used kernel in remote sensing applications is the Gaussian RBF kernel:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad (12)$$

where γ denotes the width of the Gaussian RBF.

Besides the standard kernel functions, some specific kernels have been introduced for spectral-spatial classifications (Camps-Valls et al., 2006) and multisource and multitemporal applications (Camps-Valls et al., 2008).

The training process requires the estimation of the γ and the regularization parameter C . Many approaches have been introduced, which use a leave-one-out cross-validation procedure for an automatic model selection (Chapelle et al., 2001; Chung et al., 2003).

The output of a SVM classifier is different from other algorithms. Whereas classifiers such as a decision tree directly provide a final class label and the MLC results in probability measures, the SVM provides distance measures between each pixel and the separating hyperplane.

Due to the binary nature of SVM, a multiclass decomposition strategy is necessary to handle multiclass problems. Although it is possible to solve multiclass problem directly (Foody), two main strategies exist for solving multiclass problems: A c -class problem is divided into several binary problems. Following a one-against-one (OAO) strategy, classification problem with c classes is separated into $c(c-1)/2$ subproblems, and $c(c-1)/2$

SVM classifiers are trained, one for each possible pair of classes ω_i and ω_j , with $1 \leq i < j \leq c$. In contrast to this, the one-against-all strategy (OAA) is based on a set of c classifiers and separates each class ω_i from the rest $\Omega - \omega_i$. The maximum decision value $f(x)$ determines the final class membership.

In contrast to the above strategies, it is possible to transfer the original outputs into class probabilities (Lin et al., 2003; Wu et al., 2004). Thereby, probability values can be used to determine the final class membership. However, the reliability of these pseudo probabilities can be questioned (Foody, 2008; Roscher et al., 2012b) and the use of alternative probabilistic models, such as import vector machines (e.g., Roscher et al., 2012a), is interesting.

Summary

Land cover classification is perhaps the most widely used application in remote sensing image analysis. In this entry, some well-known and state-of-the-art classifier algorithms were presented. Regarding recent and upcoming Earth observation missions, the availability of remote sensing data will increase and a variety of diverse data types can be expected for the future. Due to enhanced missions with increased revisit times and better spatial resolutions, the potential of the use of these approaches is increasing further and thereby remote sensing applications will become even more attractive. Thus, the classification of remote sensing data is and will continue to be an important research topic in the field of remote sensing.

Bibliography

- Backhaus, B., and Beule, B., 2005. Efficiency evaluation of satellite data products in environmental policy. *Space Policy*, **21**, 173–183.
- Benediktsson, J. A., Swain, P. H., and Ersoy, O. K., 1990. Neural network approaches versus statistical-methods in classification of multisource remote-sensing data. *IEEE Transaction on Geoscience and Remote Sensing*, **28**, 540–552.
- Benediktsson, J. A., Sveinsson, J. R., and Amason, A., 1995. Classification and feature extraction of AVIRIS data. *IEEE Transaction on Geoscience and Remote Sensing*, **33**, 1194–1205.
- Bishop, C. M., 1995. *Neural Networks for Pattern Recognition*. Oxford: Clarendon.
- Breiman, L., Friedman, J., Olshen, R. A., and Stone, C. J., 1984. *Classification and Regression Trees*. Boca Raton, FL: CRC Press.
- Briem, G. J., Benediktsson, J. A., and Sveinsson, J. R., 2002. Multiple classifiers applied to multisource remote sensing data. *IEEE Transaction on Geoscience and Remote Sensing*, **40**, 2291–2299.
- Bruzzone, L., and Prieto, D., 1999. A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images. *IEEE Transaction on Geoscience and Remote Sensing*, **37**, 1179–1184.
- Bruzzone, L., Marconcini, M., Wegmuller, U., and Wiesmann, A., 2004. An advanced system for the automatic classification of multitemporal SAR images. *IEEE Transaction on Geoscience and Remote Sensing*, **42**, 1321–1334.
- Burges, C. J. C., 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, **2**, 121–167.
- Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Rojo-Alvarez, J. L., Martinez-Ramon, M., Vila-Frances, J., and Calpe-Maravilla, J., 2006. Composite kernels for hyperspectral image classification. *IEEE Transaction on Geoscience and Remote Sensing*, **3**, 93–97.
- Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Rojo-Alvarez, J. L., and Martinez-Ramon, M., 2008. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Transaction on Geoscience and Remote Sensing*, **46**, 1822–1835.
- Chapelle, O., Vapnik, V., Bousquet, O., and Mukherjee, S., 2001. Choosing multiple parameters for support vector machines. *Machine Learning*, **46**, 131–159.
- Chi, M., and Bruzzone, L., 2007. Semisupervised classification of hyperspectral images by svms optimized in the primal. *IEEE Transaction on Geoscience and Remote Sensing*, **45**, 1870–1880.
- Chung, K.-M., Kao, W.-C., Sun, C.-L., Wang, L.-L., and Lin, C.-J., 2003. Radius margin bounds for support vector machines with the RBF kernel. *Neural Computing*, **15**, 2643–2681.
- Duda, R. O., Hart, P. E., and Stork, D. G., 2001. *Pattern Classification*, 2nd edn. New York: Wiley.
- Esposito, F., Malerba, D., Semeraro, G., and Kay, J., 1997. A comparative analysis of methods for pruning decision trees. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **19**, 476–491.
- Foody, G. M., and Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transaction on Geoscience and Remote Sensing*, **42**, 1335–1343.
- Foody, G., 2008. RVM-based multi-class classification of remotely sensed data. *International Journal of Remote Sensing*, **29**, 1817–1823.
- Ghoggali, N., Melgani, F., and Bazi, Y., 2009. A multiobjective genetic svm approach for classification problems with limited training samples. *IEEE Transaction on Geoscience and Remote Sensing*, **47**, 1707–1718.
- Huang, C., Davis, L. S., and Townshend, J. R., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, **23**, 725–749.
- Jackson, Q., and Landgrebe, D. A., 2002. An adaptive method for combined covariance estimation and classification. *IEEE Transaction on Geoscience and Remote Sensing*, **40**, 1082–1087.
- Lin, H.-T., Lin, C.-J., and Weng, R. C., 2003. *A note on Platt's probabilistic outputs for support vector machines*. Technical report, Department of Computer Science, National Taiwan University, Available from World Wide Web: <http://www.csie.ntu.edu.tw/~cjlin/papers/plattprob.ps>.
- Melgani, F., and Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transaction on Geoscience and Remote Sensing*, **42**, 1778–1790.
- Pal, M., and Mather, P. M., 2006. Some issues in the classification of DAIS hyperspectral data. *International Journal of Remote Sensing*, **27**, 2895–2916.
- Paola, J. D., and Schowengerdt, R. A., 1995. A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transaction on Geoscience and Remote Sensing*, **33**, 981–996.
- Peter, N., 2004. The use of remote sensing to support the application of multilateral environmental agreements. *Space Policy*, **20**, 189–195.
- Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C., and Trianni, G., 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, **113**(1), S110–S122.

- Richards, J. A., and Jia, X., 2006. *Remote Sensing Digital Image Analysis: An Introduction*, 4th edn. New York: Springer.
- Rosenqvist, A., Milne, A., Lucas, R., Imhoff, M., and Dobson, C., 2003. A review of remote sensing technology in support of the Kyoto Protocol. *Environmental Science and Policy*, **6**, 441–455.
- Roscher, R., Förstner, W., and Waske, B., 2012a. I²VM: incremental import vector machines. *Image and Vision Computing*, **30**, 263–278.
- Roscher, R., Waske, B., and Förstner, W., 2012b. Incremental import vector machines for classifying hyperspectral data. *IEEE Transaction on Geoscience and Remote Sensing*, **50**, 3463–3473.
- Safavian, S. R., and Landgrebe, D. A., 1991. A survey of decision tree classifier methodology. *IEEE Transaction on Systems, Man, and Cybernetics*, **21**, 660–674.
- Shahshahani, B. M., and Landgrebe, D. A., 1994. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. *IEEE Transaction on Geoscience and Remote Sensing*, **32**, 1087–1095.
- Vapnik, V., 1999. *The Nature of Statistical Learning Theory*. New York: Springer.
- Waske, B., and Benediktsson, J. A., 2014. Decision fusion for classification of multisource remote sensing data. In Farr, T., and Njoku, E. G. (eds.), *Encyclopedia of Remote Sensing*. New York: Springer, in press.
- Waske, B., and Benediktsson, J. A., 2007. Fusion of support vector machines for classification of multisensor data. *IEEE Transaction on Geoscience and Remote Sensing*, **45**, 3858–3866.
- Waske, B., and Braun, M., 2009. Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**, 450–457.
- Waske, B., van der Linden, S., Benediktsson, J. A., Rabe, A., and Hostert, P., 2010. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, **48**, 2880–2889.
- Wu, T.-F., Lin, C.-J., and Weng, R. C., 2004. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, **5**, 975–1005.

POLAR ICE DYNAMICS

James Maslanik
Department of Aerospace Engineering Sciences,
University of Colorado, CCAR, Boulder, CO, USA

Synonyms

Glacier motion; Ice drift; Ice rheology; Ice sheet motion; Ice transport; Sea ice motion

Definition

Polar ice. Ice cover found in the polar regions, with “polar” typically defined as locations north or south of the Arctic and Antarctic Circles, respectively (latitudes of 66.56083 °N and 66.56083 °S). The definition includes ice that forms on the oceans (sea ice), glaciers, and ice sheets.

Ice dynamics. Motion of ice cover, including large-scale patterns of ice transport, fine-scale movement of ice floes, and the interactions of ice and forcings such as winds and ocean currents.

Introduction

To many, the “polar ice caps” are viewed as areas of permanent, relatively static, and stable ice cover. In fact, all areas of ice on the oceans and land move at different rates, ranging from the relatively fast transport of sea ice of typically between 5 and 20 km/day (e.g., Emery et al., 1997; Heil and Allison, 1999) to the movement of glaciers and ice sheets (as much as 30 m/day for the fastest advancing glaciers (Joughin et al., 2004) to 1 m/year for inland ice sheets). The dynamics of sea ice is driven by the combination of winds, ocean currents, and, to a lesser degree, tides and ocean tilt (McNutt and Overland, 2003; Leppäranta, 2004; Feltham, 2008). These forcings can vary substantially over periods of hours to days. Studies of ice floe interactions, such as the production of open water areas within the ice pack (leads and polynyas), require remote sensing observations at daily or even hourly intervals, while investigations of seasonal variability in large-scale patterns of sea ice drift involve mapping ice motion on weekly or monthly scales. In contrast, the accumulated mass in glaciers and ice sheets advances under the influence of gravity, with the rate of movement controlled by slope, friction with underlying or adjacent rock, and by the mechanical properties of the ice, and require measurements over years.

There are two main aspects to the study of ice dynamics using remote sensing: direct measurement of ice displacement over time (ice motion mapping) and the analysis of specific features of the ice cover (geophysical classification) such as sea ice ridges, ice leads, crevasses, and glacier calving associated with ice motion. The former approach describes patterns and speed of ice motion, while the latter approach provides information on geophysical characteristics of the ice, such as strength, thickness, and age, without direct measurement of displacement.

The basis for observing ice movement using remote sensing methods involves the detection of the position of particular ice surface features or characteristics through comparison of images acquired at different times. The goal is to detect the position of specific surface features or general ice characteristics in each image and then to calculate the displacement of the features from one image to the next. The expected rate of ice movement, the particular characteristics of the features or characteristics to be tracked, and the spatial resolution and geolocation accuracy of the remotely sensed data determine the required image spatial resolution, the time difference between images, and the particular analysis methods that are applied. In contrast, analysis of geophysical characteristics of specific ice features does not necessarily require sequential imagery. The discussion here focuses on remote

sensing methods to measure ice displacement. Examples of methods to study geophysical ice features in remotely sensed data are described by Carsey (1992) and Lubin and Massom (2006).

Applicable data types

Any remotely sensed data that are capable of depicting ice surface features and, for ice motion mapping, cover the same area with sufficient elapsed time between data acquisitions can be used to study ice dynamics. For investigations of sea ice dynamics, visible- and thermal-band imagery from medium-resolution sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR) (Ninnis et al., 1986; Maslanik and Barry, 1990; Meier et al., 1997) provides a good combination of spatial and temporal resolution for ice motion detection. However, due to the limiting effects of frequent cloud cover and the long polar night, microwave imagery is the most common data type used for sea ice dynamics studies. Microwave data typically used include synthetic aperture radar (SAR) imagery (Fily and Rothrock, 1987; Carsey, 1992; Haarpaintner et al., 2000; Kwok et al., 2003), passive microwave imagery (Agnew et al., 1997; Emery et al., 1997; Kwok, 2008), and radar scatterometer data (Haarpaintner, 2006; Haarpaintner and Spreen, 2007). Due to the slower movement of glaciers and ice sheets, less frequent temporal sampling is required; so higher-resolution imagery such as that provided by Landsat and ASTER are effective for measuring ice movement (Kaab et al., 2005; Dowdeswell and Benham, 2006). If sufficiently geolocated, any other data types that can be used to detect surface features, such as laser altimeter products (Abdalati and Krabill, 1999), can be used to map displacement of ice features. In turn, ice motion information can be used to calculate additional geophysical information such as ice age (e.g., Maslanik et al., 2007).

The most commonly used approaches for ice motion detection are feature-based tracking, correlation-based tracking, and flow measurement using interferometry.

Feature tracking

Feature tracking involves detecting distinct surface features present in the ice cover and then following the displacement of these features from one sequential image to the next. It is particularly effective compared to correlation methods when some aspect of the features, such as rotation of ice floes or opening and closing of crevasses, tends to change rapidly relative to the time spacing between images. A variety of digital preprocessing steps including spatial resolution enhancement, filtering and edge detection to accentuate features, image thresholding, and classification can help simplify the detection of features from one image to the next. A range of feature-detection methods has been applied to ice mapping. Commonly used approaches include wavelet mapping (Liu et al., 1999; Zhao et al., 2002; Zhao and Liu, 2007) and suites

of hierarchical steps to identify individual features based on radiometric or geometric characteristics (e.g., Kwok et al., 1990; Luckman et al., 2007). Other methods such as optical flow analysis (Gutierrez and Long, 2003) can be effective but are less commonly used.

For individual case studies and situations where relatively few images are involved, visual interpretation is an effective approach for mapping ice dynamics (e.g., Lucchitta and Ferguson, 1986). The interpreter can apply a variety of image enhancement techniques such as edge detection and noise removal to highlight features, switch rapidly back and forth between images to visually highlight change, and manually adjust geolocation of each image using features that are fixed in location between the two image dates.

Image correlation

This approach, also referred to as hierarchical correlation or maximum cross-correlation, involves determining the spatial correlation between segments of sequential images. It can be considered a method of feature tracking (Pritchard et al., 2005) but is better viewed as a means of tracking aggregate characteristics of combinations of surface features. The approach is commonly applied to sea ice tracking using passive microwave and SAR imagery (Fily and Rothrock, 1987; Holt et al., 1992; Emery et al., 1997; Agnew et al., 1997; Kwok et al., 1998; Thomas et al., 2004; Kwok, 2008) and scatterometer data (Liu and Cavalieri, 1998; Zhao et al., 2002; Haarpaintner, 2006) but has also been used to map glacier motion (e.g., Bindenschadler and Scambos, 1991; Scambos et al., 1992). The latter application is sometimes referred to as speckle tracking when SAR imagery is used (e.g., Joughin et al., 2004), since it compares correlations in patterns of SAR backscatter speckle.

Interferometry

Lateral displacement of ice surfaces can also be mapped using interferometric methods applied to phase information available in sequential SAR imagery combined with elevation data (Goldstein et al., 1993; Kwok and Fahnestock, 1996; Smith, 2002; Dietrich et al., 2007). Interferometric SAR (inSAR) methods are most commonly applied to glaciers and ice sheets (e.g., Joughin et al., 1996; Rignot et al., 2004; Rignot and Kanagaratnam, 2006) but are also applicable to sea ice in some cases, such as for the study of subtle changes in relatively stable shore-fast ice that would otherwise be too small in scale to be detected using feature extraction or correlation methods (Dammert et al., 1998).

Summary

Applications of remote sensing methods to the study of ice dynamics are widespread and highly effective. Techniques used range from tracking of individual features in the ice surface to use of interferometry to extract patterns of ice transport. Nearly every type of imagery, as well as some

non-imaging data such as laser altimetry, has been applied to the important task of observing and understanding the dynamics of sea ice, glaciers, and ice sheets.

Bibliography

- Abdalati, W., and Krabill, W. B., 1999. Calculation of ice velocities in the Jakobshavn Isbrae area using airborne laser altimetry. *Remote Sensing of Environment*, **67**, 194–204.
- Agnew, T., Le, H., and Hirose, T., 1997. Estimation of large scale sea ice motion from SSM/I 85.5 GHz imagery. *Annals of Glaciology*, **24**, 305–311.
- Bindschadler, R. A., and Scambos, T. A., 1991. Satellite-image-derived velocity field of an Antarctic ice stream. *Science*, **252**, 242–252.
- Carsey, F. (ed.), 1992. *Microwave Remote Sensing of Sea Ice*. Washington, DC: American Geophysical Union, pp. 47–71.
- Dammert, P. B. G., Lepparanta, M., and Askne, J., 1998. SAR interferometry over Baltic Sea ice. *International Journal of Remote Sensing*, **19**(16), 3019–3037, doi:10.1080/014311698214163.
- Dietrich, R., Metzger, R., Korth, W., and Perlt, J., 2007. Combined use of field observations and SAR interferometry to study ice dynamics and mass balance in Dronning Maud Land, Antarctic. *Polar Research*, **18**(2), 291–298.
- Dowdeswell, J. A., and Benham, T. J., 2006. A surge of Perseibreen, Svalbard, examined using aerial photography and ASTER high resolution satellite imagery. *Polar Research*, **22**(2), 373–383.
- Emery, W. J., Fowler, C. W., and Maslanik, J. A., 1997. Satellite derived Arctic and Antarctic sea ice motions: 1988–1994. *Geophysical Research Letters*, **24**(8), 897–900.
- Feltham, D. L., 2008. Sea ice rheology. *Annual Review of Fluid Mechanics*, doi:10.1146/annurev.fluid.40.111406.1021514091-4112.
- Fily, M., and Rothrock, D. A., 1987. Sea ice tracking by nested correlations. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-25**(5), 570–580.
- Goldstein, R. M., Engelhardt, H., Kamb, B., and Frolich, R. M., 1993. Satellite radar interferometry for monitoring ice sheet motion: application to an Antarctic ice stream. *Science*, **262**, 1525–1530.
- Gutierrez, S., and Long, D. G., 2003. Optical flow and scale-space theory applied to sea-ice motion estimation in Antarctica. In *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings 2003 IEEE International*, Toulouse, Vol. 4, pp. 2805–2807.
- Haarpaintner, J., 2006. Arctic-wide operational sea ice drift from enhanced-resolution QuikSCAT/SeaWinds scatterometry and its validation. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(1), 102–107.
- Haarpaintner, J., and Spreen, G., 2007. Use of enhanced-resolution QuikSCAT/SeaWinds data for operational ice services and climate research: sea ice edge, type, concentration, and drift. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(10), 3131–3137.
- Haarpaintner, J., Kergomard, C., Gascard, J.-C., and Haugen, P. M., 2000. Sea ice dynamics observed by ERS-2 SAR imagery and ARGOS buoys in Storfjorden, Svalbard. In *Geoscience and Remote Sensing Symposium, 2000. Proceedings IGARSS 2000. IEEE 2000 International*, Hawaii, Vol. 2, pp. 467–469.
- Heil, P., and Allison, I., 1999. The pattern and variability of Antarctic sea-ice drift in the Indian Ocean and western Pacific sectors. *Journal of Geophysical Research*, **104**(C7), 15789–15802.
- Holt, B., Rothrock, D., and Kwok, R., 1992. Determination of sea ice motion from satellite images. In Carsey, F. (ed.), *Microwave Remote Sensing of Sea Ice*. Washington, DC: American Geophysical Union. Geophysical Monograph, Vol. 68, pp. 343–354.
- Joughin, I., Kwok, R., and Fahnestock, M., 1996. Estimation of ice-sheet motion using satellite radar interferometry – method and error analysis with application to Humboldt glacier, Greenland. *Journal of Glaciology*, **42**(142), 564–575.
- Joughin, I., Abdalati, W., and Fahnestock, M., 2004. Large fluctuations in speed on Greenland's Jakobshavn Isbrae glacier. *Nature*, **432**, 608–610.
- Kaab, A., Lefauconnier, B., and Melvold, K., 2005. Flow field of Kronebreen, Svalbard, using repeated Landsat 7 and ASTER data. *Annals of Glaciology*, **42**, 7–13.
- Kwok, R., 2008. Summer sea ice motion from the 18 GHz channel of AMSR-E and the exchange of sea ice between the Pacific and Atlantic sectors. *Geophysical Research Letters*, **35**, L03504, doi:10.1029/2007GL032692.
- Kwok, R., and Fahnestock, M. A., 1996. Ice sheet motion and topography from radar interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(1), 189–200.
- Kwok, R., Curlander, J. C., McConnell, R., and Pang, S. S., 1990. An ice-motion tracking system at the Alaska SAR facility. *IEEE Journal of Oceanic Engineering*, **15**(1), 44–53.
- Kwok, R., Schweiger, A., Rothrock, D. A., Pang, S., and Kottmeier, C., 1998. Sea ice motion from satellite passive microwave imagery assessed with ERS SAR and buoy motions. *Journal of Geophysical Research*, **103**(C4), 8191–8214.
- Kwok, R., Cunningham, G. F., and Hibler, W. D., III, 2003. Subdaily ice motion and deformation from RADARSAT observations. *Geophysical Research Letters*, **30**(23), 2218, doi:10.1029/2003GL018723.
- Lepparanta, M., 2004. *The Drift of Sea Ice*. Berlin: Springer, p. 266. 9783540408819. ISBN 3540408819.
- Liu, A., and Cavalieri, D. J., 1998. On sea ice drift from the wavelet analysis of DMSP SSM/I data. *International Journal of Remote Sensing*, **19**(7), 1415–1423.
- Liu, A. K., Zhao, Y., and Wu, S. Y., 1999. Arctic sea ice drift from wavelet analysis of NSCAT and special sensor microwave imager data. *Journal of Geophysical Research*, **104**(C5), 11,529–11,538.
- Lubin, D., and Massom, R., 2006. *Polar Remote Sensing: Atmosphere and Oceans*. Berlin: Springer, p. 756.
- Lucchitta, B. K., and Ferguson, H. M., 1986. Antarctica: measuring glacier velocity from satellite images. *Science*, **234**(4780), 1105–1108.
- Luckman, A., Quincey, D., and Bevan, S., 2007. The potential of satellite radar interferometry and feature tracking for monitoring flow rates of Himalayan glaciers. *Remote Sensing of Environment*, **111**, 172–181.
- Maslanik, J. A., and Barry, R. G., 1990. Remote sensing in Antarctica and the southern ocean: applications and developments. *Antarctic Science*, **2**(2), 105–121.
- Maslanik, J. A., Fowler, C., Stroeve, J., Drobot, S., Zwally, J., Yi, D., and Emery, W., 2007. A younger, thinner Arctic ice cover: increased potential for rapid, extensive sea-ice loss. *Geophysical Research Letters*, **34**, L24501, doi:10.1029/2007GL032043.
- McNutt, S. L., and Overland, J. E., 2003. Understanding the spatial hierarchy in Arctic sea ice dynamics. *Tellus A*, **55**(2), 181–191.
- Meier, W. N., Maslanik, J. A., Key, J. R., and Fowler, C. W. W., 1997. Multiparameter AVHRR-derived products for Arctic climate studies. *Earth Interactions*, **1**(5), 1–29.
- Ninnis, R. N., Emery, W. J., and Collins, M. J., 1986. Automated extraction of sea ice motion from AVHRR imagery. *Journal of Geophysical Research*, **91**, 10725–10734.
- Pritchard, H. T., Murray, A., Luckman, T. S., and Barr, S., 2005. Glacier surge dynamics of Sortebrae, east Greenland, from synthetic aperture radar feature tracking. *Journal of Geophysical Research*, **110**, F03005, doi:10.1029/2004F000233.

- Rignot, E., and Kanagaratnam, P., 2006. Changes in the velocity structure of the Greenland ice sheet. *Science*, **311**(5763), 986–990.
- Rignot, E., Braaten, D., Gogineni, S. P., Krabill, W. B., and McConnell, J. R., 2004. Rapid ice discharge from southeast Greenland glaciers. *Geophysical Research Letters*, **31**(10), 1–4.
- Scambos, T. A., Dutkiewicz, M. J., Wilson, J. C., and Bindshadler, R. A., 1992. Application of image cross-correlation to the measurement of glacier velocity using satellite image data. *Remote Sensing of Environment*, **42**, 177–186.
- Smith, L. C., 2002. Emerging applications of interferometric synthetic aperture radar (InSAR) in geomorphology and hydrology. *Annals of the Association of American Geographers*, **92**(3), 385–398.
- Thomas, M., Geiger, C., and Kambhamettu, C., 2004. Discontinuous non-rigid motion analysis of sea ice using C-band synthetic aperture radar satellite imagery. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW'04*. Washington, DC: IEEE Computer Society, pp. 1063–6919/04.
- Zhao, Y., and Liu, A. K., 2007. Arctic sea-ice motion and its relation to pressure field. *Journal of Oceanography*, **63**(3), 505–515, doi:10.1007/s10872-007-0045-2.
- Zhao, Y., Liu, A. K., and Long, D. G., 2002. Validation of sea ice motion from QuikSCAT with those from SSM/I and buoy. *IEEE Transactions on Geoscience and Remote Sensing*, **40**(6), 1241–1246.

POLAR OCEAN NAVIGATION

Lawson Brigham
University of Alaska, Fairbanks, AK, USA

Synonyms

Ice navigation; Polar ship routing; Sea ice monitoring

Definition

Polar ocean navigation. Ship navigation through waters in the Arctic and Southern oceans covered with sea ice and icebergs.

Introduction

Modern remote sensing is critically important to ship navigation in all polar seas. With its broad array of sensors and applications, the field of remote sensing has revolutionized the ability of ships to effectively and safely navigate in ice-covered waters. Expanded use of remote sensing technologies has also enhanced polar marine safety and marine environmental protection. The application of remote sensing to ship navigation began with the development of radar. Early marine radars from the Second World War provided evidence of their ability to detect ice edges, small ice floes, and icebergs. As shipboard radars advanced, sea ice features could be detected several nautical miles from the ship, although a key limitation was that the return of sea ice could be masked by the normal sea return (the high-intensity echoes or clutter) on many systems. Small sea ice features could be lost in the sea return and escape detection; however, larger icebergs and ice

floes could generally be detected by most shipboard radar systems. Enhanced radar processing technology (producing high-quality, low-intensity radar video images) has recently been developed to greatly improve conventional marine radar for use in ice navigation.

The potential ice-monitoring capability of meteorological satellites was realized with the launch of the US satellite TIROS I in April 1960. The early visible sensors were limited to daylight and cloud-free conditions, and therefore, useful ice information from visible images for ship operations was infrequent. Since the polar regions are characterized by long periods of winter darkness and cloudy conditions, researchers recognized that microwave observing systems might be more advantageous since microwaves can penetrate clouds and are independent of light conditions (Rees, 2006). Satellite microwave sensors were developed by the Soviet Union and United States, and these were flown in the late 1960s and early 1970s, opening a new era of synoptic and all-weather observations for polar navigation. Satellite passive microwave radiometers have become the backbone of global sea ice monitoring since the late 1970s, providing daily information on sea ice coverage in the Arctic Ocean and around Antarctica. This large-scale ice information has been incorporated in sea ice charts developed at the national ice centers which pass their ice products to marine operators in both polar regions. Active microwave systems – real-aperture side-looking radar (SLR) and synthetic aperture radar (SAR) – were used during the 1970s to the early 1990s for aircraft ice monitoring of Arctic regions with marine traffic (Falkingham, 1991). Satellite SLR systems were also used by the Soviet Union to monitor sea ice along the Northern Sea Route during the 1980s and 1990s (Alexandrov et al., 2004; Johannessen et al., 2007). However, it has been the successful development of satellite SAR in the 1990s that has provided the ice centers and polar navigators with high-resolution ice imagery. A series of satellites with SAR sensors – ERS-1 (1991) and ERS-2 (1995) of the European Remote Sensing program, the Canadian RADARSAT-1 (1995) and RADARSAT-2 (2007), and the European ENVISAT (2002) – have provided high-resolution SAR images that have enhanced ice charting and onboard ice information (Bertoia et al., 1998; Tsatsoulis and Kwok, 1998).

Levels of ice information for polar navigation

Ship navigation through ice in polar seas generally requires two basic levels of information aboard ship: tactical, or information required for immediate navigation such as course alteration and maneuvering, and strategic, or information required for daily track planning and course decision-making. Tactical and strategic ice information can come from the ship's radar systems and from satellite data (passive microwave data, not SAR) which can be transmitted and received directly aboard ship in near real time (within minutes of the satellite passing the ship). A third level of information can be considered operational

planning which includes information of both short and long time scales. This type of information, heavily dependent on remotely sensed imagery, is typically developed ashore in ice or weather centers that provide polar mariners with forecasts and generate predictions for ice conditions (days, weeks, and season ahead of transits), weather conditions, transit difficulty, and even optimized routes. These three categories of ice information needs for navigation in both polar regions are broadly defined, and there can be considerable overlap of sea ice spatial scales and remote sensor performances among the three.

Key sea ice and climatological parameters that are critical to ice navigation and drive sensor requirements include the following:

- Topography (ice ridges, rafts, leads): tactical and strategic information
- Thickness: tactical, strategic, and operational planning
- Concentration: tactical and strategic
- Ice Floe Size and Small-Scale Features: tactical
- Snow Cover: tactical
- Strength (ice age, first- and multi-year ice): tactical, strategic, and operational planning
- Ice Drift: strategic and operational planning
- Regional Climatology (especially wind velocities): operational planning

Each of these parameters has varying spatial and temporal scales (Hutchings and Bitz, 2005). Thus, several remote sensing instruments may be required to provide adequate information about a particular sea ice characteristic. The challenge for the shipboard ice navigators and the shore-based ice centers is to combine an integrated strategy of remotely sensed observations (across multiple spatial and temporal scales) with ice forecasts that are based on sea ice models.

Observing systems

Table 1 is a summary of the general categories of observing systems used early in the twenty-first century for ice

navigation. All are based on the remote sensing capabilities of satellite sensors, airborne sensors, and shipboard radars. Included are selected key characteristics on coverage, image resolution, scale, and delivery issues. Data transmission remains one of the key challenges or constraints on remote sensing information used for ice navigation. The ultimate timeliness of sea ice images and product analyses is crucial to their effective use in tactical navigation. Low-resolution satellite images, both passive microwave and optical (visible and infrared), have been received directly aboard icebreakers and other polar ships since the mid-1980s. Onboard receivers track polar orbiting satellites and download multiple, daily (and twice daily) images from passive microwave and optical sensors; most of this information (12–25 km spatial resolution) has limited use for tactical or immediate navigation, although it is of significant value for planning and regional understanding of weather and sea ice conditions. In contrast, high-resolution raw data (from active microwave sensors such as SAR) – with high information content to 10 m resolution including ice concentration, ice topography, and ice types – must be received by ground stations in the high latitudes and then passed to ice centers for processing. The ice centers then transmit ice analysis products and processed imagery to polar ships.

During 1991–2005 Norwegian and Russian researchers of the Nansen Center conducted a series of pioneering demonstrations applying the use of satellite SAR images for summer and winter navigation along Russia's Northern Sea Route. ERS-1/ERS-2, RADARSAT-1, and ENVISAT images were successfully used to assist winter convoys (Alexandrov et al., 2000; Johannessen et al., 1997), including a major expedition (ARCDEV) exploring the potential for year-round tanker traffic from the Ob River to Europe (Pettersson et al., 1999). The SAR images were used onboard by icebreaker navigators to select favorable sailing routes based on the high-resolution ice imagery; the images could also be inserted into the ship's electronic chart display system.

Polar Ocean Navigation, Table 1 Selected observing systems for ice navigation

Observing system	Coverage/image resolution	Scale/delivery issues
Satellite systems: low resolution	Large-scale: 10–50 km	Large-scale information Daily delivery
Passive microwave and scatterometers	Sea ice edge, concentration, ice drift	Real-time delivery aboard ship No thickness data
Satellite systems: high resolution Synthetic aperture radar (SAR), optical and infrared (IR)	Regional and local scale: 50 km down to 10 m Ice floes, surface roughness ridges	Small-scale information Near-real-time delivery from ice and weather centers Ice thickness information limited
Airborne systems: high resolution SAR, LIDAR, and electromagnetic induction	Higher resolution Ice types, ice floes, ice roughness, ridges (height, freeboard, thickness)	Small-scale information along flight paths Near-real-time delivery
Shipboard radar	Local sea ice and iceberg coverage near ship	Small-scale information Real-time information

Role of the ice centers and commercial providers

Polar ocean navigation and remote sensing are closely linked through the national ice centers. These centers (such as the Canadian Ice Service, Russia's Arctic and Antarctic Research Institute, the Norwegian Meteorological Institute, the US National Ice Center, and other members of the International Ice Charting Working Group or IICWG) provide sea ice analyses and forecast products to a host of marine users operating in the Arctic and Antarctic. Most sea ice and weather information a polar ship mariner receives has been produced at an ice center or regional meteorology office, and most of these products are based on satellite imagery (from passive and active sensors) provided to the ice centers. Ice products are also based on aerial and shipboard observations, in situ sensors, and the use of computer modeling; however, satellite remote sensing remains the best means of observing remote, polar areas with extensive sea ice coverage. The operational ice service centers and the users are generally concerned with four requirements: the ice features to be measured, the spatial scale of the observations, the frequency of the observations, and the delivery time between the observations and when useful sea ice information will be received aboard ship.

Commercial providers have also developed services to provide satellite images and other products for tactical use by polar ship operators. For example, a commercial Norwegian enterprise, Kongsberg Satellite Services or KSAT (jointly owned by the government Norwegian Space Centre and the private Kongsberg Defence & Aerospace), can provide high-resolution satellite images of sea ice to an Internet connection or satellite phone anywhere in the world. KSAT can send European ENVISAT and Canadian RADARSAT SAR images which contain detailed and time-sensitive sea ice information that can be used for planning the safe and most effective route for a ship through the existing ice conditions. Commercial services will support polar ocean navigation more routinely in the future as more marine users require near-real-time and high-resolution imagery that may potentially include better sea ice thickness information.

Future challenges and requirements

Satellite remote sensing of sea ice and ocean climatology will play increasing key roles in polar ship navigation, especially in the Arctic Ocean where marine access is improving with the marked retreat of summertime sea ice. Increased repeat satellite coverage and data fusion (to fill spatial-temporal gaps) will be required so that higher-frequency ice information can be achieved. The greatest challenge, and perhaps the most pressing need by ice navigators, is for enhanced sea ice thickness measurements by satellite and airborne remote sensors. The CryoSat-2 satellite launched in 2010 by the European Space Agency carries an advanced radar altimeter capable of higher-resolution measurements of sea ice

thickness. The altimeter measures the freeboard of the ice, or the height of the ice above the ocean surface. Improved regional and local sea ice thickness maps will necessitate continued research into improving airborne electromagnetic measurements (Wadhams and Amanatidis, 2006).

Satellite C-band SAR sensors will continue to be improved resulting in the enhancement of daily and weekly Arctic sea ice charts (Johannessen et al., 2007). Higher-resolution SAR images should provide levels of detail for small sea ice features sufficient for ice centers to provide future charts with transit difficulty and route optimization information on a regional scale. The use of expert systems and other artificial intelligence technologies will assist in the improvement of sea ice features for multiple-scale, remotely sensed images. Current and future improvements for shipboard marine radars include requirements for enhanced radar image processing and high-resolution display systems. Real-time, tactical sea ice and iceberg information at the local scale should improve. In summary, the use of remote sensing in ship navigation throughout Arctic and Antarctic waters during the twenty-first century will be pervasive and ultimately indispensable to safe and efficient polar marine operations.

Bibliography

- Alexandrov, V., Sandven, S., Johannessen, O., et al., 2000. Winter navigation in the northern sea route using RADARSAT data. *Polar Record*, **36**, 335–344.
- Alexandrov, V., Sandven, S., Kloster, K., et al., 2004. Comparison of sea ice signatures of OKEAN and RADARSAT radar images for the northeastern Barents Sea. *Canadian Journal of Remote Sensing*, **30**, 883–893.
- Bertoia, C., Falkingham, J., and Fetterer, F., 1998. Polar SAR data for operational sea ice mapping. In Tsatsoulis, C., and Kwok, R. (eds.), *Analysis of SAR Data of the Polar Oceans*. Berlin: Springer.
- Falkingham, J., 1991. Operational remote sensing of Sea Ice. *Arctic*, **44**(Suppl. 1), 29–33.
- Hutchings, J., and Bitz, C., 2005. *Sea Ice Mass Budget of the Arctic (SIMBA) Workshop: Bridging Regional and Global Scales*. University of Alaska Fairbanks.
- Johannessen, O., Sandven, S., Kloster, K., et al., 1997. ERS-1/2 monitoring of dangerous ice phenomena along the western part of northern sea route. *International Journal of Remote Sensing*, **18**, 2477–2481.
- Johannessen, O., Alexandrov, V., Frolov, I., Sandven, S., et al., 2007. *Remote Sensing of Sea Ice in the Northern Sea Route: Studies and Applications*. Chichester, UK: Springer-Praxis.
- Pettersson, L., Sandven, S., Dalen, O., et al., 1999. Satellite radar ice monitoring for ice navigation of the ARCDEV tanker convoy in the Kara Sea. In *Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions*, August 23–27, 1999, Helsinki, Finland, pp. 141–153.
- Rees, W. G., 2006. *Remote Sensing of Snow and Ice*. Boca Raton, FL: CRC Press.
- Tsatsoulis, C., and Kwok, R. (eds.), 1998. *Analysis of SAR Data of the Polar Oceans*. Berlin: Springer.
- Wadhams, P., and Amanatidis, G. (eds.), 2006. *Arctic Sea Ice Thickness: Past, Present and Future*. Brussels: European Commission.

Cross-references

Ice Sheets and Ice Volume
Icebergs
Microwave Radiometers
Radars
Radar, Synthetic Aperture
Sea Ice Concentration and Extent

Environmental Treaties
International Collaboration
Law of Remote Sensing
Mission Costs of Earth-Observing Satellites
Operational Transition
Public-Private Partnerships

POLICIES AND ECONOMICS

Roberta Balstad
CIESIN, Columbia University, Palisades, NY, USA

Definition

Satellite remote sensing systems, particularly civil Earth observation systems, whether constructed in the public or the private sector, are generally governed by national policies. There are several reasons for this. Civil remote sensing satellites can be used to obtain observations in numerous areas of national importance, such as weather, climate, natural hazards, environment, and even international relations. As a consequence, control and management of these satellite systems is critically important to governments. This control is exerted through legislative, regulatory, and fiscal policies.

A second reason that governments are involved in policies for remote sensing is that these systems are extremely expensive to build and to operate. In most cases, governments initially became involved in remote sensing because of their need for specific types of remotely sensed data and their desire for technological expertise in this area. Over time, in their search for ways to defray some of the costs of building and operating satellites, many governments began to collaborate with private sector organizations under various types of public-private sector partnerships and government-to-government partnerships. These partnerships can involve the development, construction, or operation of remote sensing systems and generally also cover the release of the data. In most public/private sector partnerships, the government partner provides financial support to the private sector partner in exchange for participation in setting performance requirements and the retention of rights to the data under specified conditions. The private sector partner generally retains the right to commercialize the data after government requirements have been met.

In this section of the encyclopedia, entries discuss a broad range of economic, social, legal, and [data access](#) issues that, together with scientific and technical capacity, have contributed to shaping the satellite remote sensing systems in place today.

Cross-references

Commercial Remote Sensing
Cost Benefit Assessment
Data Policies

PRECISION AGRICULTURE

Kelly Thorp
USDA-ARS U.S. Arid-Land Agricultural Research
Center, Maricopa, AZ, USA

Synonyms

Farming by soils; Farming by the foot; Precision crop management; Precision farming; Site-specific agriculture; Site-specific management

Definition

Precision agriculture. A modern agricultural management paradigm that utilizes information technology to tailor crop management strategies according to inherent spatial and temporal variability within agricultural systems.

Information technology. Use of computer-based systems to collect, store, process, transmit, retrieve, and protect information.

Management zone. An area within an agricultural field that has a similar characteristic or that requires a similar treatment strategy for fertilizer or pesticide.

Decision support system. A software system that utilizes information about an agricultural system to predict the outcomes of various management alternatives and to aid in the decision-making process for agricultural crop management.

Precision management strategy. Any agricultural crop management decision that is made within a precision agriculture context.

Introduction

Before the advancement of agricultural mechanization in the previous century, the land area that an individual farmer could manage was relatively small. Without the aid of modern agricultural machinery, these earlier farmers relied on their experience and intuition to understand the spatial and temporal variability of soil characteristics and crop responses across the landscape, such that labor and other resources necessary for agricultural production could be most efficiently utilized. However, with agricultural mechanization came the ability of an individual farmer to manage larger areas of land, and farm management based on inherent soil and crop growth variability became secondary to the goal of increasing production by uniformly managing agricultural crops over larger land units (Morgan and Ess, 1997). Precision agriculture is a modern crop management paradigm that proposes to merge these dissimilar management objectives by utilizing information technology on

the farm. With agricultural computerization, farmers are able to use agricultural machines for managing large land units while accounting for inherent soil and crop spatial and temporal variability at the same time (Plant, 2001; Cox, 2002; Robert, 2002).

Spatial and temporal variability

Key to the concept of precision agriculture is the idea that many different aspects of agricultural systems, including soil properties, soil moisture, crop growth and development, crop stress levels, crop yield, crop residues, and pest infestations, are variable in space and time. If this variability can be appropriately measured, analyzed, and interpreted, then that information can be used to facilitate and improve the management decisions that ultimately govern the performance of agricultural systems. Pursuit of such an endeavor has potential to make our agricultural practices more productive, profitable, sustainable, and environmentally responsible (Zhang et al., 2002).

Remote sensing for precision agriculture

Remote sensing has been proposed as a viable source of information for understanding variability within agricultural systems, particularly in the spatial dimension (Frazier et al., 1997; Moran et al., 1997). Remote sensing images of appropriate spatial resolution have been useful for delineating management zones, which represent areas of the field that have similar characteristics or that require similar treatment strategies for pest infestations, crop stress, crop disease, soil tilth, or soil water deficit. Because of its raster nature, image-based remote sensing permits the division of an agricultural scene according to the spectral variability within the image, such that certain areas of the field can be managed differently from others.

Developments in remote sensing technology have improved our ability to monitor crop growth, development, and yield and to assess the effects of physical and biological stresses on crop productivity (Hatfield and Pinter, 1993). Within a precision agriculture context, these remote sensing techniques are applied with the objective to understand how various crop characteristics change in space and time, such that crop management can be adjusted accordingly. For example, since plant nitrogen contents have been linked to canopy reflectance measurements in the green, red, and near-infrared portions of the electromagnetic spectrum, remote sensing has been useful for delineation of field areas that have a nitrogen deficiency and require additional fertilizer. Infrared remote sensing has been used to quantify crop water stress, and infrared thermometers are currently being tested as a decision aid for precision water management with center-pivot irrigation systems. Use of remote sensing to detect abnormal canopy responses or to identify different plant species offers potential for delineation of weed infestations, such that herbicides can be applied only where necessary. Remote sensing indices can also be used to locate areas of abnormal crop growth caused by crop disease or insect infestations. Remote sensing-based

assessments of crop growth and development throughout the growing season are useful for monitoring the overall vegetative vigor of the plant canopy and for identifying the time at which the crop enters resource-sensitive developmental stages. In addition to being a useful aid for spatial and temporal crop management decisions, such assessments have also shown promise for forecasting and mapping the final crop yield.

Other precision agriculture objectives can be addressed by collecting remote sensing data over fallow fields (Moran et al., 1997). Spectral reflectance measurements over bare soil surfaces have been related to variation in soil physical and chemical properties, including soil nutrient contents, soil organic matter contents, soil salinity, soil texture, and soil compaction. Variability in these soil characteristics has also been directly related to variability in crop productivity. As a result, remote sensing has been proposed as a tool for mapping soil variability with the idea that zones for precision crop management could be established based on these soil maps. Variability in crop productivity is also highly dependent on soil moisture. Various types of instruments have been used to monitor soil moisture variability, including synthetic aperture radar and passive sensors for reflectance in the microwave, infrared, near-infrared, visible, and gamma portions of the electromagnetic spectrum. In the context of precision agriculture, remote sensing has been proposed as a tool for understanding the influence of topography on field-level soil moisture regimes. In addition, remote sensing of soil moisture has been suggested as a means to map the location of artificial subsurface drainage lines that were installed prior to the existence of modern mapping techniques. Spectral reflectance from bare soils is also affected by soil surface roughness and surface residue cover, which makes remote sensing a valuable tool for precision crop residue management. On a regional scale, remote sensing has been proposed as a tool for mapping variation in tillage practices, such that farmer compliance with erosion control guidelines could be verified.

Current obstacles

Robert (2002) reasoned that widespread adoption of precision agronomic practices is currently limited by socioeconomic, agronomical, and technological obstacles. Socioeconomic obstacles involve the cost of implementation and a lack of computer and technological skills among the majority of farmers worldwide. Agronomic barriers include the lack of basic information on variability within soil and crop systems, misuse of the information that is available, lack of qualified agronomic consulting services, and lack of a methodology for generating accurate site-specific management recommendations. Further technological developments in farm machinery, instrumentation, global positioning, computer software, and remote sensing are also necessary.

Several factors currently limit the usage of remote sensing technology in precision agriculture

(Moran et al., 1997; Robert, 2002). Some of these limitations, including cloud obstruction, partial canopy reflectance issues, instrument calibration needs, and data preprocessing requirements, are common in many applications of remote sensing. Other limiting factors are related specifically to the implementation of remote sensing within the precision agriculture context. For example, further investigations are needed to identify the precision agriculture objectives for which remote sensing is best suited and to understand the spatial, spectral, and temporal image resolution requirements for effectively addressing these objectives. Standardized methodologies for proper analysis and interpretation of remote sensing images are also necessary to insure that appropriate strategies for precision crop management are developed and disseminated in a timely fashion. Finally, there must be trained specialists in rural areas to coordinate remote sensing data collection, processing, and interpretation.

Importance of unified technologies

The technical success of precision agriculture will ultimately depend on the coordinated utilization of several different technological tools for designing and implementing precision management strategies. Required technologies include remote sensing, geographic information systems, spatial statistics, computer simulation models, global positioning equipment, and machine-based sensors, instrumentation, and control systems. Within this framework, the strength of remote sensing is its ability to rapidly and cost-effectively capture the spatial variability of reflectance from agricultural fields, which can subsequently be used to estimate the spatial variability of several important crop and soil characteristics. However, translation of spectral reflectance and spectral indices into justifiable management decisions has proven to be more difficult. One proposed solution has been to couple remote sensing with crop growth simulation modeling, where remote sensing would provide spatial information for model calibration using data assimilation strategies. Model simulations within a precision agriculture decision support framework would then be used to derive strategies for precision management of agricultural systems.

Conclusion

Precision agriculture is a modern agricultural management paradigm that promises increased agricultural productivity, profitability, sustainability, and environmental responsibility by utilizing information technology in every aspect of farm management. Remote sensing is an important component of the precision agriculture concept, because it can be used to rapidly estimate the inherent spatial and temporal crop and soil variability that exists within all agricultural systems. Continued advancements in precision agriculture technologies will help to improve agricultural production efficiencies, promote natural resource conservation, and secure the production of our food, feed, fiber, and fuel needs for many generations in the future.

Bibliography

- Cox, S., 2002. Information technology: the global key to precision agriculture and sustainability. *Computers and Electronics in Agriculture*, **36**, 93–111.
- Frazier, B. E., Walters, C. S., and Perry, E. M., 1997. Role of remote sensing in site-specific management. In Pierce, F. J., and Sadler, E. J. (eds.), *The State of Site Specific Management for Agriculture*. Madison, WI: ASA-CSSA-SSSA, pp. 149–160.
- Hatfield, J. L., and Pinter, P. J., Jr., 1993. Remote sensing for crop protection. *Crop Protection*, **12**, 403–413.
- Moran, M. S., Inoue, Y., and Barnes, E. M., 1997. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment*, **61**, 319–346.
- Morgan, M., and Ess, D., 1997. *The Precision-Farming Guide for Agriculturalists*. Moline, IL: Deere & Company.
- Plant, R. E., 2001. Site-specific management: the application of information technology to crop production. *Computers and Electronics in Agriculture*, **30**, 9–29.
- Robert, P. C., 2002. Precision agriculture: a challenge for crop nutrition management. *Plant and Soil*, **247**, 143–149.
- Zhang, N., Wang, M., and Wang, N., 2002. Precision agriculture – a worldwide overview. *Computers and Electronics in Agriculture*, **36**, 113–132.

Cross-references

[Agriculture and Remote Sensing](#)
[Commercial Remote Sensing](#)
[Crop Stress](#)
[Data Assimilation](#)
[Irrigation Management](#)
[Soil Moisture](#)
[Soil Properties](#)
[Vegetation Indices](#)
[Vegetation Phenology](#)

PROCESSING LEVELS

Ron Weaver
 National Snow and Ice Data Center, Cooperative
 Institute for Research in Environmental Sciences,
 University of Colorado, Boulder, CO, USA

Definition

Processing levels. A means of describing the way remote sensing digital data are processed from raw or engineering units to informational geophysical products. These are typically differentiated by numeric (or sometimes alphabetical) hierarchies, from Level Zero, indicating engineering units or the data from the actual sensor system, to Level Four indicating geophysical products. Typically processing levels are applied to satellite data streams, but they have been utilized in other forms of remote sensing data.

Introduction

Processing levels as used in remote sensing activities are a methodology for differentiating raw (low) from more highly processed forms of digital data, typically collected

Processing Levels, Table 1 Processing levels as defined by CODMAC

Definitions of space science data levels and types

Data level or type	Definition	Utility
1. Raw data	Telemetry data with data embedded	Little use to most of science community, except for radio sciences
2. Edited data	Corrected for telemetry errors and split or decommuted into a data set for a given instrument. Sometimes called Experimental Data Record. Data are also tagged with time and location of acquisition. Corresponds to NASA Level 0 data	Wide use, especially for researchers familiar with instrumentation
3. Calibrated data	Edited data that are still in units produced by instruments but that have been corrected so that values are expressed in or are proportional to some physical unit such as radiance. No resampling, so edited data can be reconstructed. NASA Level 1A	
4. Resampled data	Data that have been resampled in the time or space domains in such a way that the original edited data cannot be reconstructed. Could be calibrated in addition to being resampled. NASA L1B	
5. Derived data	Derived results, as maps, reports, graphs, etc. NASA Levels 2 through 5	General way in which information is transferred
6. Ancillary data	Nonscience data needed to generate calibrated or resampled data sets. Consists of instrument gains, offsets; pointing information for scan platforms, etc.	Needed to be able to convert edited data to calibrated, resampled, or derived data sets
7. Correlative data	Other science data needed to interpret spaceborne data sets. May include ground-based data observations such as soil type or ocean buoy measurements of wind drift	Crucial data in many cases to provide ground truth calibration for spaceborne data
8. User description	Description of why the data were acquired, any peculiarities associated with the data sets, and enough documentation to allow secondary user to extract information from the data	Important aspect associated with the data that will be even more important for facility-class instruments and for secondary users of data

by satellite systems. The practice of defining processing levels originates from the space physics research community early in the 1970s civilian satellite research era. First references to processing levels within the US remote sensing community are traced to the Committee on Data Management and Computation (CODMAC) of the National Research Council (NRC, 1982, 1986). These reports codified the practices within NASA and NOAA, which predate the CODMAC effort (see Table 1). The CODMAC hoped that, by providing standard definitions, semantic confusion within the research community would be alleviated. This in fact has come to pass. Today, a search of the Web for “Data Processing Levels” reveals many current programs within NASA, NOAA, and other agencies which employ these definitions (Kobler, 1995; EOS, 2006; Chandra, 2008).

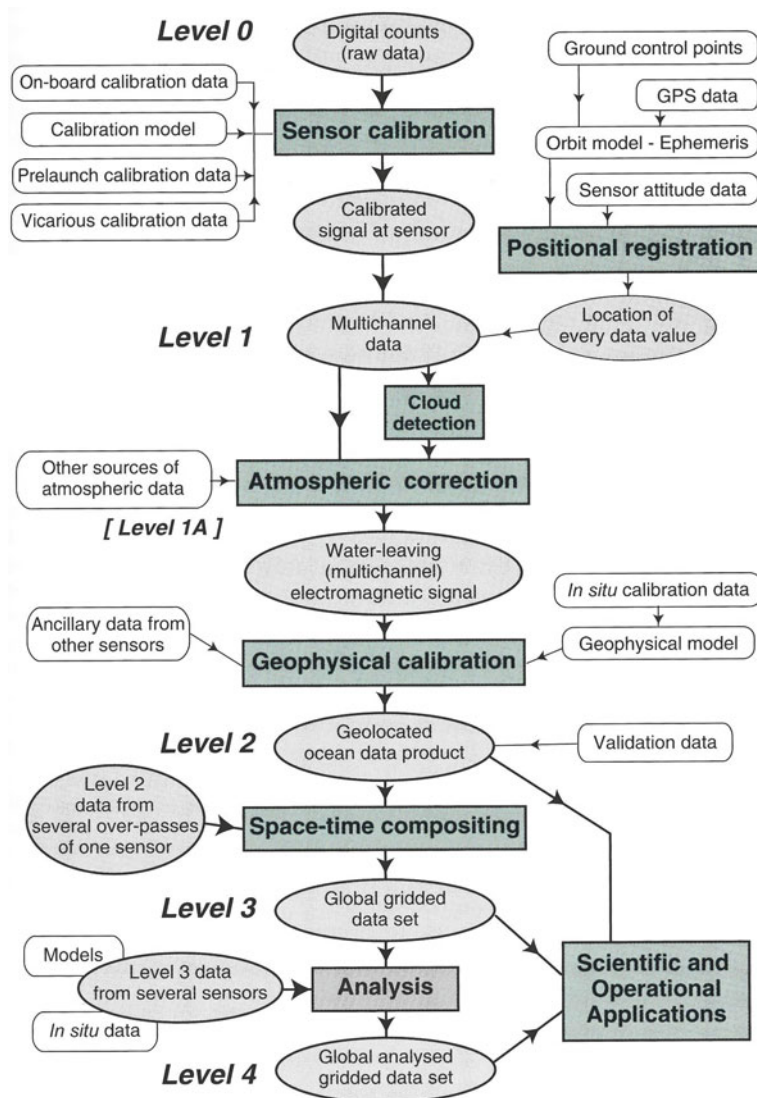
Other systems have been devised throughout the satellite era. For example, the World Climate Program (WCRP) follows a system originally devised for the Global Atmospheric Research Program (GARP) in the 1970s. GARP divided processing into three levels or groups: (1) primary instruments, (2) retrieved geophysical parameters, and (3) analysis fields. These categories roughly match the NASA/CODMAC levels 0–3. This WCRP classification has been used as is or in slightly modified form by WCRP programs such as Tropical Ocean and

Processing Levels, Table 2 Data processing levels from NASA Earth observing system

Data processing level definitions

Processing level	definition
Level 0	Reconstructed, unprocessed data at full resolution; all communications artifacts have been removed
Level 1	Level 0 data that has been time-referenced and annotated with ancillary information, including radiometric and geometric calibration coefficients, and geolocation information
Level 2	Derived geophysical variables at the same resolution and location as the Level 1 data
Level 3	Variables mapped on uniform space-time grids, usually with some completeness and consistency
Level 4	Model output or results from analyses of lower-level data

Global Atmosphere (TOGA), Global Energy and Water Cycle Experiment (GEWEX), or the International Satellite Cloud Climatology Project (ISCCP) (Gutman et al., 1995, 1997).



Processing Levels, Figure 1 Schematic data processing flow (Robinson, 2004).

It is important to differentiate from processing levels as defined in this section and levels of service (LOS). The data management community is increasingly active in the process of defining levels of service for data sets, which help prioritize resource allocations to existing data sets (Hunolt, 2003). Higher LOS means a data set might be housed in a more responsive data system or with more extensive metadata. LOS is complementary to processing levels but is clearly not the same. In the Internet-web serving environment, another related term has emerged. Quality of Service (QoS), or the ability to prioritize web information delivery by a predefined set of rules, has become a necessary part of data delivery systems and implementation of LOS for data.

Defined processing levels

The definitions presented here follow NASA practices. Table 2 presents the generally accepted definitions in use today by multiple NASA and NOAA missions. A more descriptive work flow diagram was developed by Robinson (2004, p. 99) and is presented in Figure 1. Others have identified processing steps and these contribute to processing level descriptions. For example, Kobler (1995) lays out the following science data processing requirements:

- Time ordering of data and removal of duplicate packets
- Repackaging of satellite instrument data with additional spacecraft and engineering data
- Calibration of data to correct for instrument drift

- Conversion of data to environmental variables
- Resampling of data to convert to gridded Earth patterns
- Generation of higher-level science analysis and modeling products

These steps generally follow the processing levels of Table 2.

Implementations of processing levels

Since the early 1970s, NOAA polar orbiter data has been labeled with processing levels. Those familiar with the NOAA data streams will recognize the L1A and L1B designations of AVHRR data. Level 1A (see Figure 1) refers to processing raw telemetry or L0 data to a time-sequenced but not decommutated format. Typically L1A data remain internal to NOAA archives. L1A processing is generally reversible to L0. Level 1B are discrete instrument-specific data sets derived from L1A and retain full resolution but are time-referenced and annotated with ancillary information including data quality indicators, calibration coefficients, and georeferencing parameters (NOAA, 2001).

NASA Earth remote sensing satellites have utilized the processing level concepts for many missions. The Earth Observing System (EOS) from its inception in the early 1990s to present designates data sets and products using processing levels (EOS, 1992).

Outside of the US National efforts by NOAA and NASA, the Committee on Earth Observing Satellites (CEOS) has developed a similar processing level descriptive hierarchy. The CEOS classifies data products by how they are processed and secondarily by their expected use (Gutman et al., 1997).

Summary

Use of processing levels in descriptions of remote sensing data and data processing over the past 40 plus years has been helpful to both data producers and data users. Many variations of level definitions exist internationally, with most deriving from the early usage within the NASA and NOAA remote sensing communities. In the United States remote sensing community, the most commonly applied definitions may be found on the NASA Earth Observing System documentation (EOS, 2006).

Bibliography

- Chandra, (2008). Standard data processing (webpage). Available from World Wide Web: <http://cxc.harvard.edu/ciao/data/sdp.html>
- Committee on Data Management and Computation (CODMAC), 1982. In Bernstein, R., et al. (eds.), *Data Management and Computation Volume 1: Issues and Recommendations*. Washington, DC: National Academy Press.
- Committee on Data Management and Computation (CODMAC), and National Research Council, 1986. *Issues and Recommendations Associated with Distributed Computation and Data Management Systems for the Space Sciences*. Washington, DC: National Academy Press.

Earth Observing System (EOS), 2006. In Parkinson, C. (ed.), *Earth Science Reference Handbook*. Greenbelt, MD: Goddard Space Flight Center, p. 277. Available from World Wide Web: http://eospsoc.gsfc.nasa.gov/ftp_docs/2006ReferenceHandbook.pdf

Gutman, G., and Alexander, I., 1997. Towards a common language in satellite data management: a new processing level nomenclature. In *Proceedings Geoscience and Remote Sensing, 1997. IGARSS'97. Remote Sensing - A Scientific Vision for Sustainable Development*. Piscataway, NJ: IEEE International, Vol. 3, pp. 1252-1254.

Gutman, G., Tarpley, D., Ignatov, A., and Olson, S., 1995. The enhanced NOAA global land dataset from the advance very high resolution radiometer. *Bulletin of the American Meteorological Society*, 76(7), 1141-1156.

Hunolt, G., (2003). SEEDS levels of service/cost estimation study. In *Working Paper Five: Data/Provider Model, Requests and Levels of Service*. NASA Goddard Space Flight Center, p. 29. Available from World Wide Web: http://esdswg.gsfc.nasa.gov/st_cost.html#Documents.

Kobler, B., and John, B., 1995. Architecture and design of storage and data management for the NASA earth observing system data and information system (EOSDIS). In *Fourteenth IEEE Symposium on Mass Storage Systems*. Washington, DC: IEEE Computer Society, pp. 65-76, 1051-9173/95.

NOAA, (2001). NOAA KLM Users Guide, Section 6.0 Ingest and Pre-processing. Available from World Wide Web: <http://www.ncdc.noaa.gov/oa/pod-guide/ncdc/docs/klm/index.htm>

Cross-references

[Data Access](#)

[Data Archival and Distribution](#)

PUBLIC-PRIVATE PARTNERSHIPS

William Gail

Global Weather Corporation, Boulder, CO, USA

Definition

Partnership. A formal relationship or association between two or more organizations established to pursue business opportunities.

Public-private. A collaboration between a government entity and a private sector company.

Introduction

Public-private partnerships are a formal relationship between a government organization and one or more private sector companies. They are typically formed to pursue a specific objective of interest to the government entity for which the private sector partners present an efficient means of accomplishing that objective and involve a contractual relationship that defines the roles and responsibilities of each partner along with the ownership share. The relationship often takes the form of a single-purpose joint venture, sometimes referred to as a special-purpose vehicle (SPV). The government

may provide direct support such as one-time capital contributions or indirect support such as tax subsidies. Ideally, a public-private partnership provides goods and services at a higher quality and lower cost than either the government or private sector could do alone.

The most common examples of public-private partnerships occur in infrastructure development. These include roadways, hydroelectric projects, airports, and similar facilities. Development of pharmaceutical products for otherwise neglected illnesses in the developing world is also sometimes accomplished through public-private partnerships, enabling greater availability of vaccines and other medical resources than might otherwise be achieved through traditional for-profit markets.

General history and issues

Public-private partnerships have been in use for more than a century. An early example is the Scottish railway developer Robert Gillespie Reid who built numerous railway lines and bridges for both the USA and Canadian Governments in the late 1800s. His partnerships specified that his company would build railway facilities and operate them for 50 years, at which time it would own them. As a result, he was able to turn limited capital investment into significant land rights.

The advantage of such partnerships to the government is the ability to leverage skills and resources of the business sector while retaining a strong say in how an activity is pursued. The advantage to the private sector entity is the financial opportunity that arises from a relatively stable, long-term business.

A common criticism of public-private partnerships is that the government partner bears the primary risk while the private sector partner accrues most of the financial benefit. For example, financing risks such as interest rates and international exchange rates represent a risk that can be borne by either the government or the private sector entity. A well-defined partnership carefully addresses this balance between risk and reward to the benefit of all involved parties.

Partnerships and remote sensing policy

Within the remote sensing sector, public-private partnerships have been used by various nations to build, deploy, and operate Earth observation systems. In most cases, these systems are intended for dual-use purposes, such as joint military and civilian applications. Commercial sales of imagery are often an important objective; they may represent only a small element of the data use or can be the system's primary purpose. Within these partnerships, it has become increasingly difficult to distinguish between public and private entities (Gabrynowicz, 2007). In part, nations have different definitions of "commercial" or "private," and organizations are often linked in subtle ways through subsidies and other means. In Europe, for example, "commercial" generally refers to any entity that generates

revenue, whether privately owned or state owned. State-owned organizations are sometimes even organized similar to and operate like private corporations.

Public-private partnerships are governed by a combination of policy and law. Nations such as the USA and Canada have transparent policies and laws concerning commercialization of remote sensing. Others address the issue through less-open internal policies. Some have policies but no corresponding law. In nearly all instances, however, the greatest interest in partnerships occurs for high-resolution imagery, for which the overlap between commercial and national security interests is strongest and thus most policy sensitive.

In the USA, public-private partnerships for remote sensing are governed in part by the *Land Remote Sensing Policy Act of 1992*, created largely to resolved problems within the Landsat program. This act states that "commercialization of land remote sensing should remain a long-term goal of United States policy." It also notes that "development of the remote sensing market and the provision of commercial value-added services based on remote sensing data should remain exclusively the function of the private sector." One of the breakthrough provisions in this policy and its 1984 predecessor is a means for licensing fully commercial remote sensing systems: "In consultation with other appropriate United States Government agencies, the Secretary is authorized to license private sector parties to operate private remote sensing space systems for such period as the Secretary may specify and in accordance with the provisions of this subchapter."

Both public-private partnerships and commercial remote sensing are closely tied to national industrial and national security interests. In the USA, this aspect of land remote sensing policy was clarified in the 1994 Presidential Decision Directive 23, which stated that "the fundamental goal of our policy is to support and to enhance US industrial competitiveness in the field of remote sensing space capabilities while at the same time protecting US national security and foreign policy interests. Success in this endeavor will contribute to maintaining our critical industrial base, advancing US technology, creating economic opportunities, strengthening the US balance of payments, enhancing national influence, and promoting regional stability." This was reconfirmed by the 2003 National Security Presidential Directive 27 *US Commercial Remote Sensing Policy* stating a goal to "advance and protect US national security and foreign policy interests by maintaining the nation's leadership in remote sensing space activities, and by sustaining and enhancing the US remote sensing industry."

Outside the USA, the nature of public-private partnerships is often tied more openly to national interests, such as technology development and surveillance. The intermingling of commercial and security interests can present particular challenges, such as the need for bankers involved with investment in the partnership to hold security clearances.

Examples

Among the oldest public-private partnerships in remote sensing is for the French System pour l'Observation de la Terre (SPOT) satellite series. The Spot Image company was formed in 1982 as a public-private partnership between the French space agency Centre National d'Études Spatiales (CNES), the French national geographic center Institut Geographique National (IGN), and several French space companies to own and operate French commercial remote sensing satellite systems. As of 2010, EADS owned 99 % while IGN retained a 1 % minority interest. Spot Image has launched a series of increasingly capable satellites, starting with the 10 m resolution Spot-1 in 1986. The progression has continued with the Pleiades series of electrooptical imaging satellites launched starting 2011 and as part of a joint program called ORFEO in which Italy supplies radar satellites.

The US Landsat system has a long history of collaboration between the government and private sector entities. In 1985, the government-owned Landsat was partially privatized through a partnership with Earth Observation Satellite Corporation (EOSAT), itself a partnership of Hughes Aircraft and RCA. EOSAT was given a contract to operate the system for 10 years and build follow-on satellites, obtaining rights for commercial sale of the data. The Landsat Data Continuity Mission, initiated in the early 2000s as the follow-on to Landsat-7, was originally envisioned as a strong public-private partnership, but has since evolved into a more traditional system acquisition with the government retaining responsibility for data processing and distribution.

One of the earlier and most well-known examples of a purposely designed public-private partnership is the Sea-viewing Wide Field-of-view Sensor (SeaWiFS). It was built for NASA under partnership with Orbital Sciences Corporation and launched in 1997. ORBIMAGE (now part of GeoEye), a spin-off from Orbital Sciences, was given the responsibility for operating the satellite and for selling the data, with NASA itself purchasing data.

The intelligence community has been a significant partner in development of US commercial remote sensing capability. Through large data purchase contracts such as NextView, ClearView, and EnhancedView, the National Geospatial-Intelligence Agency (NGA) has provided anchor funding for commercial remote sensing companies. NextView contracts, for example, are service-level agreements (SLA) that specify requirements for imagery purchases from satellites still to be developed. They are designed to provide a long-term government partnership commitment that enables the private sector partner to finance and develop new satellites meeting the requirements. To motivate a commercial market, the agreements are nonexclusive, meaning that the companies can sell the same imagery to other customers.

In the mid-1990s, NASA initiated a public-private partnership called LightSAR. The objective was to develop

a dual-use synthetic aperture radar (SAR) satellite that served both scientific and commercial needs. Several teams were funded to study both the system and the market opportunity, but the final bid produced no team capable of satisfying the partnership's objectives. In this case, the scientific and commercial objectives required largely incompatible system requirements, so their combination did little to reduce costs over pursuing each independently. Dual use involving scientific-quality data also raises numerous challenges regarding open-data practices, calibration, metadata, and more. One of the teams involved in the LightSAR study included the German company Dornier as a major partner. Dornier (now EADS Astrium) subsequently built on the experience gained to form the German public-private partnership InfoTerra and launch the series of TerraSAR satellites.

Since the late 1980s, Canada has been building a public-private partnership for synthetic aperture radar (SAR) through the Radarsat series. The company Radarsat International was formed in 1989 by a consortium of Canadian companies with the purpose of processing and distributing the data from Radarsat-1, launched in 1995. One of the shareholders, Macdonald Dettwiler and Associates, bought out the others in 1999 and formed a deeper partnership with the Canadian Government to build and operate Radarsat-2, launched in 2007.

Many other nations utilize public-private partnerships for remote sensing systems. The Israeli EROS imagery satellites are owned and operated by an international company ImageSat International. Surrey Satellite Technology Ltd. (SSTL) developed a multigovernment collaboration that built and launched a series of small satellites called the Disaster Monitoring Constellation (DMC), each owned and controlled by a different government (Algeria, Nigeria, Turkey, Britain, China, Spain) and operated jointly through a wholly owned subsidiary of SSTL called DMC International Imaging. The German RapidEye satellite constellation, a five-satellite system designed for rapid refresh imagery, was developed with participation from the European Union and the German state of Brandenburg. In addition to space systems, aerial remote sensing operations can also leverage public-private partnerships.

Summary

Public-private partnerships are used extensively in remote sensing to develop and operate satellite systems. They are typically formed or supported to achieve a government-driven objective (such as surveillance or disaster monitoring) most efficiently accomplished through a private sector entity.

Bibliography

Broadbent, J., and Laughlin, R., 2003. Public-private partnerships: an introduction. *Journal of Accounting, Auditing & Accountability*, 16(3), 332–341.

- Gabrynowicz, J., 2007. *The Land Remote Sensing Laws and Policies of National Governments: A Global Survey*. Report for U. S. Department of Commerce/National Oceanic and Atmospheric Administration's Satellite and Information Service Commercial Remote Sensing Licensing Program.
- Gerrard, M., 2001. Public-private partnerships. *Finance and Development*, **38**(3), 48–51.
- Group on Earth Observations, 2005. *Global Earth Observations System of Systems GEOSS: 10-Year Implementation Plan Reference Document*. Noordwijk: ESA Publications Division.
- Moszoro, M., and Gasiorowski, P., 2008. *Optimal Capital Structure of Public-Private Partnerships*. IMF Working Paper 1/2008. Washington, DC: International Monetary Fund.

- National Research Council, 2002. *Toward New Partnerships in Remote Sensing: Government, the Private Sector, and Earth Science Research*. Washington, DC: The National Academies Press.
- Vaillancourt-Rosenau, P., 2000. *Public-Private Policy Partnerships*. Cambridge: MIT Press.

Cross-references

[Commercial Remote Sensing](#)
[Cost Benefit Assessment](#)
[Emerging Applications](#)
[Observational Systems, Satellite](#)
[Policies and Economics](#)
[Radar, Synthetic Aperture](#)

R

RADAR, ALTIMETERS

Keith Raney
Applied Physics Laboratory, Johns Hopkins University,
Laurel, MD, USA

Synonyms

Active sensor; Bathymetry; Oceanic geoid; Sea-surface height; Significant wave height

Definition

It is commonly understood that an altimeter is a down-looking radar device designed to measure relative altitude, thus to establish the height of the radar above the surface below. Although similar in form and function, the purpose of a remote-sensing altimeter is different, since usually it is mounted on a spacecraft whose orbital height is known. Remote-sensing altimeters are designed to retrieve certain geophysical parameters about the surface from the timing and shape of the radar reflections (waveforms).

Introduction

Since the first flight of an experimental altimeter aboard Skylab (1973), this class of radars has been developed primarily for observation of the oceans from Earth-orbiting satellites. Their data are comprised of a nearly continuous stream of waveforms, each one of which is the time history of the power backscattered from the surface below (thus from nadir) following each pulse transmitted by the radar. Over the ocean, sea-surface height, mean wave height, and surface wind speed can be retrieved, after averaging a sufficient number of waveforms to achieve the desired precision of the measurement (see “[Radars](#)”). These instruments usually are left “on” over their entire orbit, so that applications can be investigated regarding ice sheets, inland waters, or land forms. Since altimeters look

down, the resulting waveforms from large bodies of water are relatively simple and well understood. The altimeter’s response to surface topography is more complex, although useful results have been obtained, for example, from the ice sheets of Greenland and Antarctica, and the altimeter mode on the Magellan mission to Venus (Pettengill et al., 1991). Earth-observing altimeters usually perform first-order processing on board (including “tracking” of the time delay of the surface return) so that their mean data rate is on the order of only a few 10s of kbits/s.

Geophysical measurements

In the oceanographic application, a satellite-based altimeter systematically circles the Earth, generating sea-surface height (SSH), significant wave height SWH (significant wave height is the average height [trough to crest] of the one-third largest waves), and wind speed (WS) measurements along its nadir track. These measurements accumulate, providing unique synoptic data that have revolutionized knowledge and understanding of both global and local phenomena, from El Niño to bathymetry.

The driving requirement for oceanographic applications is determination of the local sea level relative to the Earth’s geoid. The average elevation relative to the Earth’s center of mass of the mean sea surface in the absence of dynamic perturbations is due to tides, winds, and currents. The reference for this measurement – the orbital height of the spacecraft – must be determined by means other than altimetry, such as GPS, Doppler tracking, or laser range finding (Chelton et al., 2001). The intended measurement – SSH – is given by the difference between the orbit height and the altimeter’s range measurement. The two other measurements made by a radar altimeter – SWH and WS – are retrieved from the shape of the return waveform.

Sea-surface height

Sea-surface height is a function of many geophysical parameters, such as current flow, the mean temperature in a column of water, or variations in the ocean's depth. Relatively small changes (on the order of cm) in mean sea-surface height may correspond to substantial differences in the corresponding geophysical parameters. It follows that accuracy and precision (see "*Radars*") are both required for useful measurements. An altimeter's accuracy depends to first order on precision orbit determination (POD) of spacecraft height and on correction of the propagation delays suffered by the radar's round-trip waveform. Accuracy is improved by "crossover" analysis (Luthcke et al., 2003), which is exploited to remove systematic height errors by extensive comparison of height measurements derived from intersecting ascending and descending orbits, under the assumption that the true sea level has not changed between observations. The precision of an ocean-viewing altimeter is proportional to the radar's range resolution and inversely proportional to the square root of the number of statistically independent measurements (looks) combined for each data point. Typical numbers for these key parameters are 0.5 m single-pulse post-compression range resolution and 1,500 or more looks averaged per second. Ocean-viewing altimeters in general benefit from the rather large signal ($3 \text{ dB} < \sigma_0 < 20 \text{ dB}$) reflected from the water's surface, which is quasi-specular when viewed vertically (see "*Microwave Surface Scattering and Emission*").

The SSH measurement objectives of space-based altimeters can be grouped into four broad categories: large-scale dynamic sea-surface topography, dynamic mesoscale oceanic features, static mesoscale sea-surface topography, and ice-sea ice as well as continental ice sheets. In the field of oceanography, mesoscale features have scales of several hundred kilometers, as opposed to the much larger basin scale (the North Atlantic ocean, for example). Each of these measurement themes implies narrowed constraints on preferred orbit and on the top-level instrument and mission design. Satellite altimeters dedicated to determining the ocean's large-scale dynamic surface topography are characterized by absolute sea-surface height (SSH) 1 s averaged measurement accuracy on the order of centimeters along tracks of more than 1,000 km and orbits at relatively high altitude that retrace their surface tracks in a stipulated number of days. These data when averaged over long timescales provide estimates of annual sea-level rise, known from radar altimetry to an accuracy of 1 mm per year (Nerem and Mitchum, 2001).

Wave height and wind speed

To extract SWH and WS from waveform data, finely tuned algorithms have been developed and validated against in situ buoy measurements (Fu and Cazanave, 2001). For example, the TOPEX Ku-band altimeter measured SWH to within 0.5 m up to more than 5.0 m

and WS within 1.5 m/s up to more than 15 m/s. These figures correspond to averages over 1 s, or about 7 km along the subsatellite path of the altimeter's footprint, which typically is 3–5 km wide, determined by mean sea state.

Mesoscale features

Important applications at the mesoscale include mapping of geostrophic currents [whose cross-flow SSH gradient is proportional to mean current velocity (Leeuwenburgh and Stammer, 2002; Stammer and Dieterich, 1999)] and tracking eddies (whose SSH relative to the surrounding area reflects the mean temperature of the underlying water column). Rather than absolute SSH accuracy, these shorter-scale applications require precision sufficient to sustain surface slope measurement accuracy on the order of $1 \mu\text{rad}$ (1 cm sea-level change over a 10 km distance), on spatial scales of less than several 100 km.

Oceanographic versus geodetic orbits

Both large-scale topographic and mesoscale missions are aimed at dynamic features whose SSH signals are on the order of only a few tens of centimeters and less and that remain relatively constant over time intervals up to several weeks. These signals must be distinguished from variations in SSH caused by the Earth's geoid, which is essentially constant on the scale of many years, in contrast to the dynamic signals set up by currents, volumetric temperature differentials, and the like. Globally the geoidal SSH signal spans a range of more than 100 m with respect to the reference ellipsoid of the mean Earth's figure. In order to see the dynamic topographic signals, the much larger geodetic signal has to be established, which then becomes the reference for the dynamic signals. The traditional strategy to establish this geoidal reference is to restrict the altimeter's orbit to an "exact-repeat" pattern. Once the geoid has been determined along such repeat tracks, then dynamic variations from that reference can be observed. The standard tolerance on "exact" is $\pm 1 \text{ km}$ on the maximum allowable deviation of every orbit's subsatellite locus relative to the nominal surface track; this requirement was derived from the height error induced by the mean sea-surface slope of the geoid. Exact-repeat orbits are the traditional standard for oceanographic radar altimeters, which as a consequence are characterized by track-to-track separations at the equator of several hundred kilometers. Altimeters in the TOPEX orbit, for example, have repeat tracks that are spaced apart at the equator by 316 km.

Geodetic signals are expressed as two-dimensional static topographic variations on the sea surface. To map these, orbits are required that have moderate inclination and generate dense track-to-track spacing, leading to nonrepeat (or long-period-repeat) orbits. Radar altimetric data from such orbits can be processed to measure deviations in both the N–S and the E–W planes of the local gravity vector ("deflection of the vertical") which is expressed as the local two-dimensional slope of the mean

sea surface, at scales on the order of 5–300 km. These data are sufficient to map the detailed structure of the local gravity field, which can be inverted to deduce bathymetry – the ocean’s depth (Sandwell and Smith, 1997; Smith and Sandwell, 2004). The prime example of a nonrepeat geodetic mission is the first 18 months of Geosat (MacArthur et al., 1989), for which the average orbit-to-orbit spacing was about 5 km at the equator.

Near-polar ice

Radar altimetric observation of ice-sheet surface elevations imposes its own set of requirements. Observation of near-polar ice requires high-inclination nonrepeat orbits, in contrast to altimeters in traditional oceanographic orbits, or sun-synchronous systems. Accurate mapping of continental ice-sheet elevations requires further that the altimeter have robust accuracy when viewing nonzero average surface slopes in the along-track and cross-track directions. CryoSat (URL/CryoSat, 2008), the first radar altimeter devoted to the Earth’s ice sheets, incorporate a cross-track interferometer and an along-track Doppler processing mode, both of which are designed to offset the surface slope problem. The CryoSat orbit is essentially nonrepeating at an inclination of 92°.

Range measurement accuracy

Given that the orbit’s height with respect to the Earth’s center of mass has been well determined, conversion of the altimeter’s range measurement into accurate sea-surface height is threatened by many sources of potential error, which must be estimated and then offset. There are two classes of corrections required: to compensate errors imposed by electromagnetic propagation and by the sea surface itself. After the range measurement errors are corrected, the residual accuracy usually is dominated by imperfect knowledge of the radial orbit.

Propagation delay corrections

An altimeter (like any radar) actually measures round-trip delay, not distance. For useful oceanographic radar altimeter data, the deceptively simple proportionality of range to delay time must take into account the small but significant retardation of the radar’s microwaves as they propagate. The cm-level SSH accuracy required of these instruments is much smaller than the ranging errors introduced by delays through the ionosphere and the atmosphere. The ionosphere imposes delays that are a function of frequency. In practice, these can be estimated and then corrected if the altimeter measures round-trip time delay at two different frequencies (Keihm et al., 1995). The atmospheric propagation delays (Goldhirsh and Rowland, 1982) are due to two components: the dry atmosphere and water vapor. The dry atmosphere component is relatively well known and stable at long spatial scales; in practice the resulting delay is compensated by recourse to model predictions. Delays due to water vapor in the atmosphere are variable down to scales of several

hundred kilometers (and much smaller when traversing a storm front). Standard practice is to measure the integrated water vapor contribution in the vertical column below the altimeter by a microwave radiometer, for which two or three frequencies are required. The method works well over the open ocean, but breaks down near land, whose microwave radiance corrupts the atmospheric signal.

Sea-surface corrections

The mean sea-surface elevation varies in response to the mass of the atmosphere above it. The implied (inverse) barometric correction is of the order of centimeters. Further, radar reflections at vertical incidence favor the wave troughs over their crests, leading to a SSH estimation error known as sea-state bias. This bias increases as significant wave height increases. The resulting correction (Chelton et al., 2001; Zanife et al., 2003), deduced empirically after many controlled measurements and comparisons to independent SSH and SWH measurements, is a function of the altimeter’s measurement of SWH (Hayne et al., 1994).

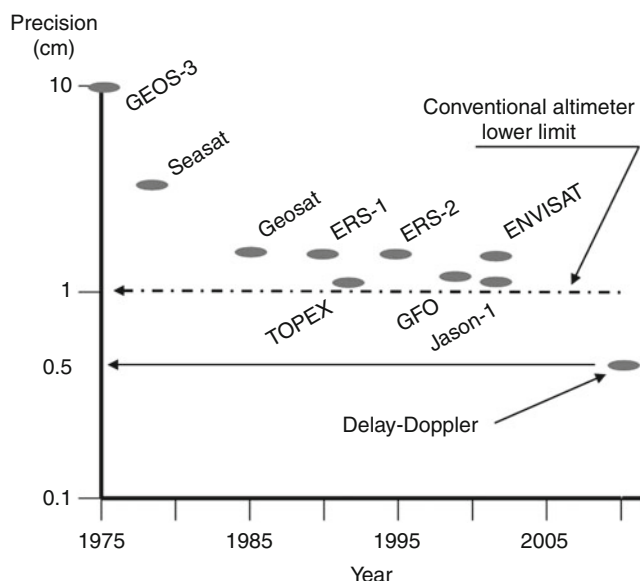
Measurement precision

Measurement precision is improved by more incoherent averaging. The lower limit usually is determined by the intrinsic precision of the radar. Figure 1 shows a summary history of these factors. The data, consistent with theory (Walsh, 1982), show that 1 cm instrument precision is close to the lower limit for conventional altimeters. The delay-Doppler instrument, which introduces waveform processing strategies adapted from synthetic aperture radar techniques (Raney, 1998) that allow increased averaging, provides approximately a factor of two improvements in precision (Jensen and Raney, 1998) to 0.5 cm.

Centimeter-scale SSH precision is supported in the oceanographic application by averaging over the range response of many returns. The resolution of each return waveform is typically on the order of 0.5 m (which is a constant of the instrument, proportional to the inverse bandwidth of the radar). As the altimeter progresses along its orbit, waveforms are accumulated and averaged pulse to pulse, whose shape converges on the flat-surface impulse response (Brown, 1977; Moore and Williams, 1957) (Figure 2). Sea-surface height is derived from the time delay to the midpoint of the waveform’s leading edge rise. About 1,500 or more such waveforms are averaged over 1 s, which corresponds to a mean range estimate whose standard deviation is on the order of 1 cm for a calm sea. In practice SSH precision degrades with increasing significant wave height.

The brown waveform

Over a quasi-flat sea, a pulse-limited altimeter’s idealized mean waveform is a modified step function, whose rise time is equal to the compressed pulse length, and whose position along the time-delay axis is determined by the altimeter’s range (Brown, 1977). If the sea surface is

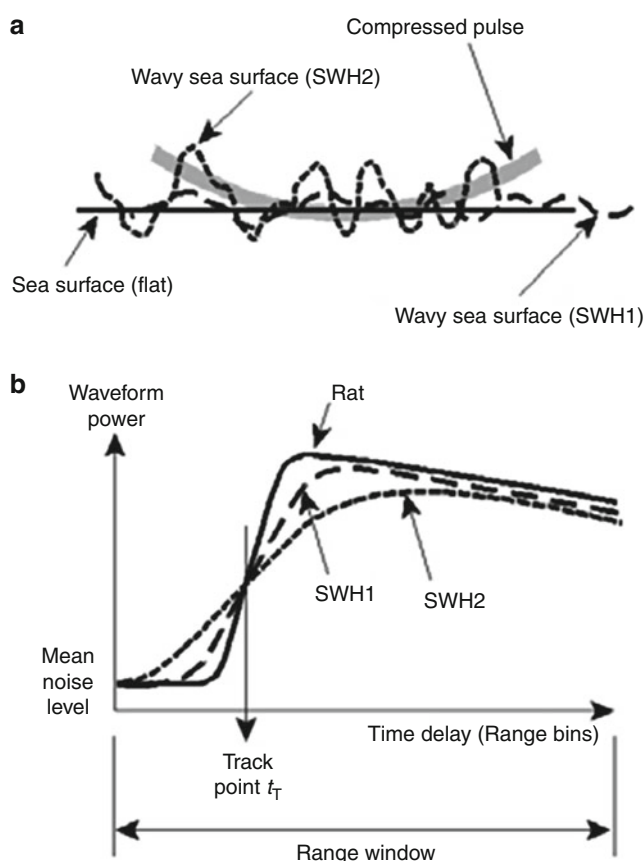


Radar, Altimeters, Figure 1 Evolution of accuracy and precision. The altimeter's precision is determined primarily by the number of statistically independent returns averaged to form each output waveform, for which the upper bound is determined by the radar's pulse-limited area. The delay-Doppler processing algorithm increases available averaging, hence improving precision by approximately a factor of 2.

modulated by waves, the altimetric depth of the surface increases, which reduces the slope of the waveform's leading edge. Hence, SWH is proportional to waveform rise time. If the sea surface is disturbed by the wind, the resulting fine-scale roughness scatters reflected energy over larger solid angles, hence decreasing the power of the reflection back to the altimeter. For wind speeds of more than about two knots, WS is inversely related to mean waveform power. In practice, the inflections of the idealized flat-surface response function waveform are softened by the pulse weighting, and the waveform is attenuated over time by the weighting of the antenna pattern and further by any mis-pointing of the antenna pattern toward nadir.

Flight systems

Selected attributes of satellite radar altimeters are summarized in Table 1. The Jason series of altimeters represent the state of the art in absolute sea-surface height measurement accuracy, due in no small measure to their high-altitude TOPEX orbits. Significant milestones since 1973 include S-193 and GEOS-3 (McGoogan et al., 1974; Hayne, 1980), proof of concept; Seasat (MacArthur et al., 1989), first full de-ramp pulse compression (full de-ramp – or simply de-ramp – is standard terminology in space-based radar altimetry; it is known more commonly to most radar engineers as the Stretch technique); *Geosat* (MacArthur et al., 1987; Sandwell and Smith, 2001; Smith and Sandwell, 1994), first geodetic mission;



Radar, Altimeters, Figure 2 Waveforms. The return waveform (averaged) of an ocean-viewing radar altimeter is a softened step function, whose position in delay time, steepness of initial rise, and peak power are processed to retrieve respectively sea-surface height, significant wave height, and surface wind speed.

TOPEX/Poseidon (Chelton et al., 2001; Goldhirsh and Rowland, 1982), first global dynamic topography mission, and the first two-frequency ionosphere correction; *ERS-1*, *ERS-2*, and *ENVISAT* (Fu and Cazanave, 2001; URL/ENVISAT, 2008; Zelli, 1999), sun-synchronous multipayload missions; *CryoSat-2* (Wingham et al., 2004), first ice-sheet radar altimeter (the original *CryoSat* was launched in October 2005, but unfortunately, the launch vehicle malfunctioned; ESA and its member states authorized a replacement); *AltiKa* (Vincent et al., 2006), first Ka-band altimeter; and *Sentinel-3* (Le Roy et al., 2007), conventional and delay-Doppler modes.

Ice-sheet radar altimetry

Pulse-limited space-based radar altimeters work best over relatively mild topographic relief of mean slope zero, such as the ocean's surface. Over ice sheets or terrestrial surfaces, their performance is degraded (beam-limited techniques, of which laser altimeters are extreme examples, circumvent these problems but imply their own set of disadvantages, including limited lifetime and the

Radar, Altimeters, Table 1 Earth-observing radar altimeters (primarily oceanographic objectives), 1973–2012

Spacecraft	Country	Year	Repeat days	Inclination degrees	Altitude km	Spacing km	Band	Accuracy cm
Skylab (3)	USA	1973	No	~48	435	n/a	Ku	50 m
GEOS-3	USA	1975–1978	No	115	845	~60	Ku	50
Seasat	USA	1978	~17, 3	108	800	160, 900	Ku	20
Geosat	USA	1985–1989	<i>GM</i> , 17.05	108	800	~5, 160	Ku	10
ERS-1	ESA	1991–1996	3, 35, 176	98.5	785	900, 80, 15	Ku	7
TOPEX/Poseidon	USA/France	1992–2005	9.916	66	1,336	315	C, Ku	2
							Ku	5
ERS-2	ESA	1995	35	98.5	785	80	Ku	7
GFO	USA	1998	17.05	108	800	160	Ku	5
Jason-1	France	2001	9.916	66	1,336	315	C, Ku	1.5
Envisat	ESA	2002	35	98.5	785	80	S, Ku	7
Jason-2	France	2008	9.916	66	1,336	315	C, Ku	1.5
Altika-3	India (Fr)	(2009)	35	98.5	785	80	Ka	1.8
CryoSat	ESA	2005 (2009)	369	92	720	n/a	Ku	(5)
Sentinel-3	Europe	(2012)	35	98.5	785	80	C, Ku	

inability to penetrate dense cloud formations). Approximately 95 % of Antarctica, for example, has slopes less than $\sim 3^\circ$, which although small is sufficient to trick a conventional altimeter into very large height errors. An unknown one-degree slope would lead to a 120 m surface height error.

The CryoSat altimeter (Phalippou et al., 2001) is the first space-based radar altimeter designed to operate over ice. Its payload instrument is the SAR/Interferometric Radar ALtimeter (SIRAL), which has three modes: conventional, SAR, and interferometric. The conventional pulse-limited mode reflects its Poseidon heritage. The SAR mode is the first space-based embodiment of the delay-Doppler architecture (Raney, 1998), which offers advantages in precision, resolution, and along-track surface slope tolerance. The interferometric mode (Jensen, 1995) is designed to measure the cross-track surface slope component. Data from each of the three modes are processed in ground-based facilities. The conventional mode is used for the open ocean (for calibration and sea-surface height reference purposes) and the central continental ice sheets that are relatively level. The interferometric mode is reserved for the more steeply sloping margins of the ice sheets. The synthetic aperture mode is used primarily over sea ice, where its sharper spatial resolution and precision support measuring the difference between the sea level and the top surface of floating ice (freeboard). Since the density of ice is relatively well known, such freeboard measurements can be inverted to estimate ice thickness (Laxon et al., 2003).

Orbit considerations

Given an arbitrarily good radar altimeter, its orbit becomes the dominant factor that may limit sea-surface height measurement accuracy (Parke et al., 1987). Earth orbits of altimeters usually revisit a given surface location only about every 10–20 days. This low revisit rate is in distinct contrast

to tides that have approximately one or two cycles per day, driven primarily by the Earth's rotation through the lunar and solar gravity fields. As a result, all tidal signals sensed by an altimeter are under sampled in the Nyquist sense. Altimetric data retain the resulting aliases, which over the course of a year or so can be identified, quantified, and calibrated out. Altimeter orbits are favored for which the tidal aliases do not fall upon signals of geophysical interest. The state of the art (with respect to SSH accuracy and large-scale circulation studies) is the Jason series, operating in the orbit originally designed for TOPEX/Poseidon. The orbit parameters include repeat period 9.9156 calendar days (unfortunately, often stated as 10 days), inclination 66° , repeat track separation at the equator 316 km, and altitude 1,336 km. The radial component of precision orbit determination (POD) was on the order of 2 cm for T/P, and Jason results show POD to a level of 1.5 cm.

The geosat orbit

The first 18 months of Geosat (1985–1989) were devoted to geodesy, hence a nonrepeating orbit at an inclination of 108° . After its geodetic mission, Geosat was controlled to maintain an exact-repeat orbit of a period of 17.0505 calendar days (sometimes inappropriately abbreviated to 17 days). From this orbit, half of the principal tidal constituents alias into unwanted frequencies (near zero, one, or two cycles per year). In particular, the dominant tidal constituent, the common twice-daily lunar tide, is aliased to 317 days, which is close to the annual cycle (Ray, 2001).

Sun-synchronous orbits

Sun-synchronous satellites host European Space Agency (ESA) altimeters including RA-1 on ENVISAT. All share the same orbit: 35.00 calendar days repeat period, 98.5° inclination, and 781 km mean equatorial altitude. Radial knowledge of these sun-synchronous orbits is good to about 5 cm, based on the Delft model

(Scharroo and Visser, 1998). As sun-synchronous altimeters, the largest solar constituent (twice-daily) aliases to zero, and all tidal constituents that are primarily dependent on solar forces alias to frequencies close to zero.

Radar altimeter design

Conventional oceanographic radar altimeters are pulse-limited, in the sense that their instantaneous footprint on the surface is determined by the shortness of the (compressed) range pulse, not by the width of the antenna pattern. (The footprint of an altimeter that has an extremely narrow radiation pattern such as a laser usually is beam-limited.)

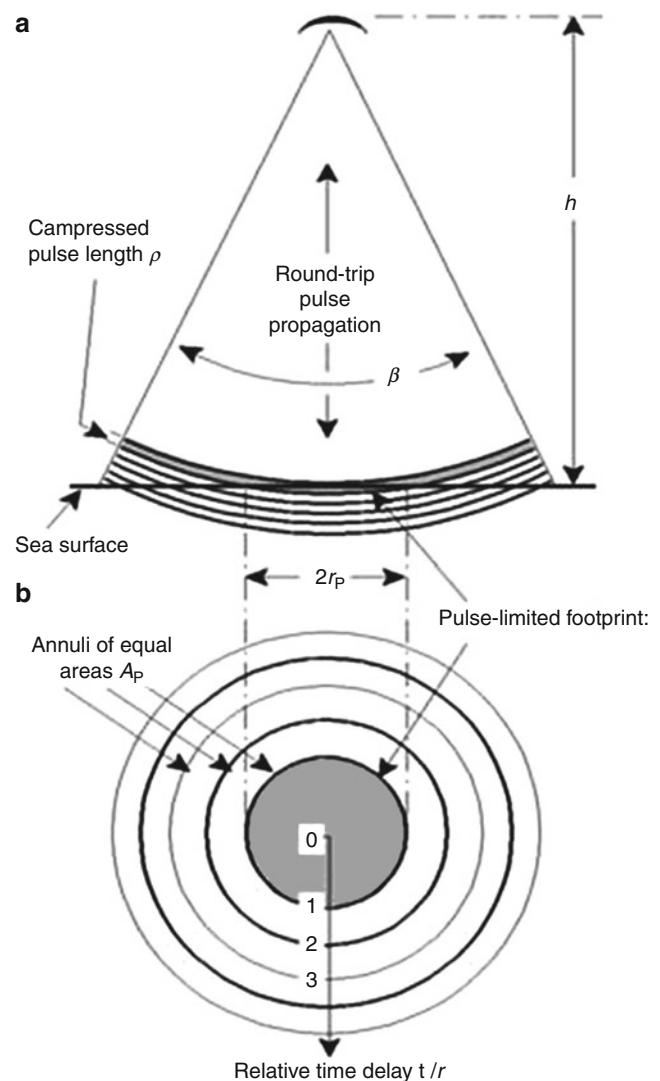
Figure 3 illustrates the pulse-limited condition (Chelton et al., 1989). The radius r_P of the area delimited on a quasi-flat surface by a pulse of (compressed) length τ seconds on the Earth of mean radius R_E seen from a relative altitude of h kilometers is

$$r_P = (c\tau h/\alpha)^{1/2} \quad (1)$$

where $\alpha = (R_E + h)/R_E$ is a consequence of the spherical observation geometry. For typical satellite radar altimeters, the pulse-limited footprint over a quasi-flat surface is on the order of 2 km in diameter. As the pulse continues to impinge and spread over the surface, the resulting pulse-limited annuli all have areas equal to that of the initial pulse-limited footprint. Hence, the received power tends to maintain the level corresponding to the peak of the initial response (Figure 2). The pulse-limited area expands in response to increasing large-scale surface roughness, which in the oceanographic context is expressed as significant wave height SWH. The height accuracy of a pulse-limited altimeter is much less sensitive to (small) angular pointing errors than is the case for a beam-limited altimeter. The classic analysis of a pulse-limited altimeter's waveform is due to Brown (1977). The waveform of a delay-Doppler altimeter (Raney, 1998) is pulse-limited in the across-track direction, but it is beam-limited in the along-track direction.

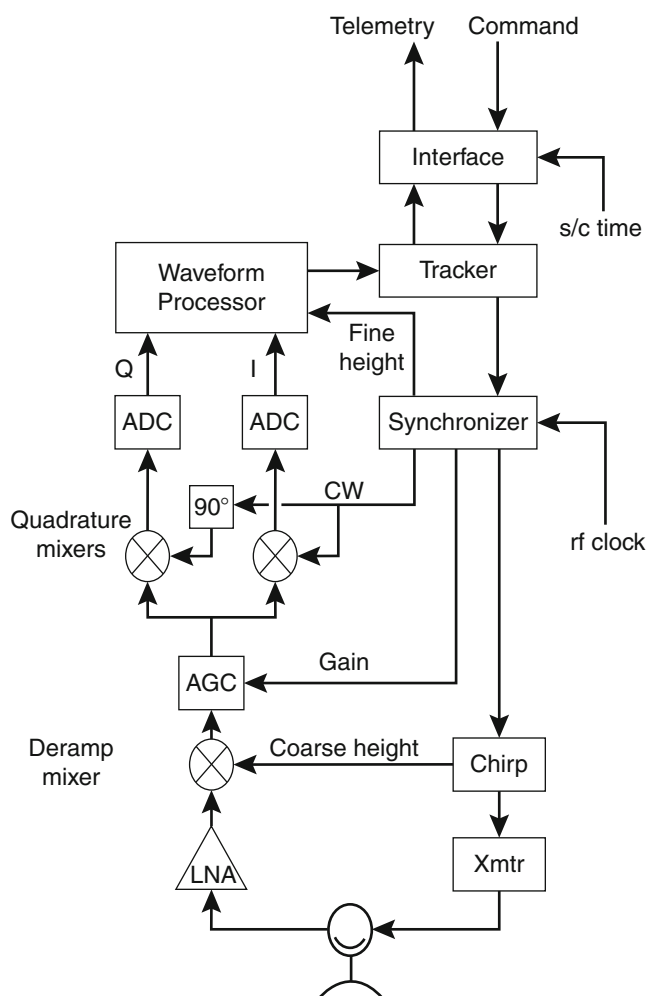
A satellite-based radar altimeter (Figure 4) needs to measure range accurately, but only for an essentially planar surface, oriented orthogonally to the radar's line of sight. Hence, ocean-viewing altimeters have a small range window whose position and gain adjusts the delay and gain of the radar (originally implemented as an alpha-beta tracker) in response to the timing and strength of the surface reflection (MacArthur et al., 1989). For certain applications it may be advantageous in subsequent processing to "retrack" the waveforms (Deng and Featherstone, 2006) for more accurate retrieval of waveform parameters.

Most satellite-based radar altimeters since Seasat use the Stretch technique (Caputi, 1971). The distinguishing feature of this method of pulse modulation and demodulation is a clever trade between the two key parameters in a large time-bandwidth product (TBP) signal



Radar, Altimeters, Figure 3 Pulse-limited geometry. In a pulse-limited radar altimeter the intersection of the compressed ranging pulse with the sea-surface determines the area of the effective footprint. The pulse-limited area increases with increasing SWH and is elongated along track in proportion to waveform averaging time.

(see “*Radars*”). After reception, demodulation is applied that transforms the “short time, large bandwidth” nature of the original pulse to “long time, small bandwidth” signals. Since the same TBP is maintained, the original resolution is preserved. The method is ideal for altimetry, since the range depth of the ocean's surface is very much smaller than the time available in the pulse repetition period. The full de-ramp technique offers a considerable savings in system bandwidth at all subsequent stages and at no cost in range resolution. The waveform bandwidth reduction can be considerable, being a factor of more than 200 for TOPEX, for example (Marth et al., 1993; Zieger et al., 1991). The data rate is reduced further by



Radar, Altimeters, Figure 4 Block diagram of the TOPEX radar altimeter, showing the essential attributes of pulse modulation and demodulation (the full de-ramp technique), range gate position and receiver gain adaptive tracking, and on-board waveform processing.

pulse-to-pulse incoherent averaging. Thus, a typical single-pulse resolution of 0.5 m, which requires a bandwidth of 300 MHz, is reduced to an average waveform data rate on the order of 10 kb/s after de-ramp demodulation, on-board tracking, and one-second averaging.

Conclusion

Radar altimeter measurements have become essential for a wide variety of applications in oceanography, geodesy, geophysics, and climatology. For retrieval algorithms and in-depth reviews of applications, the book to consult is Fu and Cazanave (2001). An excellent on-line entry point for oceanographic radar altimeter missions may be found at URL/AVISO (2008), and for geodetic applications, see URL/Geodesy (2003). A tutorial on radar altimetry may be found at URL/Altimeter_Tutorial (2008).

Although relatively mature, radar altimeters continue to evolve. Interest is increasing for extending altimetry to near-shore oceanic applications and inland waters, including surface hydrography. The constraint on repeating orbits for mesoscale oceanographic applications may be relaxed by use of the refined geoid data from GRACE and other space-based gravity measuring missions, thus opening the way for one orbit plan to satisfy both mesoscale oceanographic and geodetic requirements. Radar instrumentation is moving ahead, pioneered by the SAR mode (delay-Doppler) on CryoSat and Sentinel-3. More radical demonstrations are in store, including a Ka-band instrument. There is abiding interest in the principle of wide-swath altimetry, which would be the “perfect” instrument if the technique could extend current accuracy and precision standards to areas either side of nadir.

Bibliography

- Brown, G. S., 1977. The average impulse response of a rough surface and its applications. *IEEE Transactions on Antennas and Propagation*, **25**, 67–74.
- Caputi, W. J. J., 1971. Stretch: a time-transformation technique. *IEEE Transactions on Aerospace and Electronic Systems*, **AES-7**(2), 269–278.
- Chelton, D. B., Walsh, E. J., and MacArthur, J. L., 1989. Pulse compression and sea-level tracking in satellite altimetry. *Journal of Atmospheric and Oceanic Technology*, **6**, 407–438.
- Chelton, D. B., Ries, J. C., Haines, B. J., Fu, L.-L., and Callahan, P. S., 2001. Satellite altimetry. In Fu, L.-L., and Cazanave, A. (eds.), *Satellite Altimetry and Earth Sciences*. San Diego: Academic.
- Deng, X., and Featherstone, W. E., 2006. A coastal retracking system for satellite radar altimeter waveforms: application to ERS-2 around Australia. *Journal of Geophysical Research*, **111**, doi:10.1029/2005JC003039.
- Fu, L.-L., and Cazanave, A. (eds.), 2001. *Satellite Altimetry and the Earth Sciences*. San Diego, CA: Academic.
- Goldhirsh, J., and Rowland, J. R., 1982. A tutorial assessment of atmospheric height uncertainties for high-precision satellite altimeter missions to monitor ocean currents. *IEEE Transactions on Geoscience and Remote Sensing*, **20**, 418–434.
- Hayne, G. S., 1980. Radar altimeter mean return waveforms from near-normal incidence ocean surface scattering. *IEEE Transactions on Antennas and Propagation*, **AP-28**(5), 687–692.
- Hayne, G. S., Hancock, D. W., III, and Purdy, C. L., 1994. The corrections for significant wave height and attitude effects in the TOPEX radar altimeter. *Journal of Geophysical Research*, **99**, 24941–24955.
- Jensen, J. R., 1995. Design and performance analysis of a phase-monopulse radar altimeter for continental ice sheet monitoring. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium IGARSS'95*. Florence, Italy: IEEE.
- Jensen, J. R., and Raney, R. K., 1998. Delay Doppler radar altimeter: better measurement precision. In *Proceedings of the IEEE Geoscience and Remote Sensing Symposium IGARSS'98*. Seattle, WA: IEEE.
- Keihm, S. J., Janssen, M. A., and Ruf, C. S., 1995. TOPEX/Poseidon microwave radiometer (TMR) III: wet tropospheric range correction and pre-launch error budget. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 147–161.
- Laxon, S., Peacock, N., and Smith, D., 2003. High interannual variability of sea ice thickness in the Arctic region. *Letters to Nature*, **425**, 947–950.

- Le Roy, Y., Deschaux-Beaume, M., Mavrocordatos, C., Aquirre, M., and Heliere, F., 2007. SRAL SAR radar altimeter for Sentinel-3 mission. In *Proceedings of the International Geoscience and Remote Sensing Symposium*. Barcelona, Spain: IEEE.
- Leeuwenburgh, O., and Stammer, D., 2002. Uncertainties in altimetry-based velocity estimates. *Journal of Geophysical Research*, **107**(C10), 3175–3191.
- Luthcke, S. B., Zelensky, N. P., Rowlands, D. D., Lemoine, F. G., and Williams, T. A., 2003. The 1-centimeter orbit: Jason-1 precision orbit determination using GPS, SLR, DORIS, and altimeter data. *Marine Geodesy*, **26**(3–4), 399–421.
- MacArthur, J. L., Marth, P. C., and Wall, J. G., 1987. The GEOSAT radar altimeter. *Johns Hopkins APL Technical Digest*, **8**(2), 176–181.
- MacArthur, J. L., Kilgus, C. C., Twigg, C. A., and Brown, P. V. K., 1989. Evolution of the satellite radar altimeter. *Johns Hopkins APL Technical Digest*, **10**(4), 405–413.
- Marth, P. C., Jensen, J. R., Kilgus, C. C., Perschy, J. A., MacArthur, J. L., Hancock, D. W., Hayne, G. S., Purdy, C. L., Rossi, L. C., and Koblinsky, C. J., 1993. Prelaunch performance of the NASA altimeter for the TOPEX/Poseidon project. *IEEE Transactions on Geoscience and Remote Sensing*, **31**(2), 315–332.
- McGoogan, J. T., Miller, L. S., Brown, G. S., and Hayne, G. S., 1974. The S-193 radar altimeter experiment. *Proceedings of the IEEE*, **62**(6), 793–803.
- Moore, R. K., and Williams, C. S., Jr., 1957. Radar return at near-vertical incidence. *Proceedings of the IRE*, **45**(2), 228–238.
- Nerem, R. S., and Mitchum, G. T., 2001. Sea level change. In Fu, L.-L., and Cazenave, A. (eds.), *Satellite Altimetry and Earth Sciences*. San Diego: Academic.
- Parke, M. E., Stewart, R. H., Farless, D. L., and Cartwright, D. E., 1987. On the choice of orbits for an altimetric satellite to study ocean circulation and tides. *Journal of Geophysical Research*, **92**(C11), 11693–11707.
- Pettengill, G. H., Ford, P. G., Johnson, W. T. K., Raney, R. K., and Soderblom, L. A., 1991. Magellan: radar performance and data products. *Science*, **252**, 260–265.
- Phalippou, L., Rey, L., DeChateau-Thierry, P., Thouvenot, E., Steunou, N., Mavrocordatos, C., and Francis, R., 2001. Overview of the performances and tracking design of the SIRAL altimeter for the CryoSat mission. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Florence.
- Raney, R. K., 1998. The delay Doppler radar altimeter. *IEEE Transactions on Geoscience and Remote Sensing*, **36**(5), 1578–1588.
- Ray, R. D., 2001. Applications of high-resolution ocean topography to ocean tides. In Chelton, D. (ed.), *Report of the High-Resolution Ocean Topography Science Working Group Meeting*. Corvallis, Oregon: Oregon State University.
- Sandwell, D. T., and Smith, W. H. F., 1997. Marine gravity anomaly from Geosat and ERS-1 satellite altimetry. *Journal of Geophysical Research*, **102**, 10039–10054.
- Sandwell, D. T., and Smith, W. H. F., 2001. Bathymetric Estimation. In Fu, L.-L., and Cazenave, A. (eds.), *Satellite Altimetry and Earth Sciences*. New York: Academic.
- Scharroo, R., and Visser, P., 1998. Precise orbit determination and gravity field improvement for the ERS satellites. *Journal of Geophysical Research*, **103**(C4), 8113–8127.
- Smith, W. H. F., and Sandwell, D. T., 1994. Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry. *Journal of Geophysical Research*, **99**, 21803–21824.
- Smith, W. H. F., and Sandwell, D. T., 2004. Conventional bathymetry, bathymetry from space, and geodetic altimetry. *Oceanography*, **17**(1), 8–23.
- Stammer, D., and Dieterich, C., 1999. Space-borne measurements of the time-dependent geostrophic ocean flow field. *Journal of Atmospheric and Oceanic Technology*, **16**, 1198–1207.
- URL/Altimeter_Tutorial, 2008. Available from World Wide Web: http://earth.esa.int/brat/html/general/overview_en.html.
- URL/AVISO, 2008. Available from World Wide Web: <http://www.aviso.oceanobs.com/>.
- URL/CryoSat, 2008. Available from World Wide Web: <http://www.esa.int/export/esaLP/cryosat.html>. European Space Agency.
- URL/ENVISAT, 2008. Available from World Wide Web: <http://envisat.esa.int/>.
- URL/Geodesy, 2003. Available from World Wide Web: <http://www.ngdc.noaa.gov/mgg/bathymetry/predicted/explore.HTML>.
- Vincent, P., Steunou, N., Caubet, E., Phalippou, L., Rey, L., Thouvenot, E., and Verron, J., 2006. AltiKa: a Ka-band altimetry payload and system for operational altimetry during the GMES period. *Sensors*, **6**, 208–234.
- Walsh, E. J., 1982. Pulse-to-pulse correlation in satellite radar altimetry. *Radio Science*, **17**(4), 786–800.
- Wingham, D. J., Phalippou, L., Mavrocordatos, C., and Wallis, D., 2004. The mean echo and echo cross-product from a beamforming interferometric altimeter and their application to elevation measurements. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(10), 2305–2323.
- Zanife, O. Z., Vincent, P., Amarouche, L., Dumont, J. P., Thibaut, P., and Labroue, S., 2003. Comparison of the Ku-band range noise level and the relative sea-state bias of the Jason-1, TOPEX, and Poseidon-1 radar altimeters. *Marine Geodesy*, **26**, 201–238.
- Zelli, C., 1999. ENVISAT RA-2 Advanced radar altimeter: instrument design and pre-launch performance assessment review. *Acta Astronautica*, **44**(7–12), 323–333.
- Zieger, A. R., Hancock, D. W., Hayne, G. S., and Purdy, C. L., 1991. NASA radar altimeter for the TOPEX/Poseidon project. *Proceedings of the IEEE*, **79**(6), 810–826.

Cross-references

Cryosphere, Measurements and Applications
 Geodesy
 Ocean Surface Topography
 Ocean, Measurements and Applications
 Radars
 Sea Level Rise

RADAR, SCATTEROMETERS

David Long
 Department of Electrical and Computer Engineering,
 BYU Center for Remote Sensing, Brigham Young
 University, Provo, UT, USA

Synonyms

Microwave scatterometer; Spaceborne scatterometer;
 Wind scatterometer

Definitions

Radar scatterometer. A calibrated radar designed to measure the radar backscatter cross section of a target, generally an area on the earth's surface.

Wind scatterometer. A radar scatterometer designed to measure the ocean's surface backscatter at multiple azimuth angles in order to estimate the near-surface

vector wind. It is also used for ice melt/freeze, soil moisture, and vegetation remote sensing.

Spaceborne scatterometer. A wind scatterometer operating on an Earth-orbiting satellite. Currently operating examples: SeaWinds on QuikSCAT (USA) and ASCAT on MetOP (ESA).

Normalized radar cross section. Area-normalized radar backscatter which is the ratio of the incident and reflected radar signal power. Conceptually similar to “albedo,” but for radar backscatter.

Geophysical model function (GMF). An empirical or analytic relationship between ocean radar backscatter and wind direction and speed used for wind estimation from wind scatterometer measurements.

Signal-to-noise ratio (SNR). Ratio of the signal power (or energy) to the noise power (or energy).

Introduction

A radar scatterometer is a carefully calibrated remote-sensing radar designed to measure the radar backscatter cross section (usually as the normalized radar backscatter coefficient, denoted σ^0) of a target, which is generally a natural scene. Measurements of the backscatter coefficient are used to derive key scene properties. For example, a wind scatterometer is designed to measure backscatter cross section over the ocean from several azimuth angles. From the backscatter measurements, the near-surface wind speed and direction are estimated (retrieved).

Scatterometer fundamentals

In order to achieve high calibration accuracy, scatterometers are generally real-aperture radars. Microwave scatterometers operate by transmitting a microwave signal toward the target and measuring the reflected or backscattered power. Knowing the measurement geometry and radar parameters, the radar cross section is computed from the backscattered power. Key target characteristics, such as wind speed and direction over the ocean, are then estimated from the target cross section.

Due to thermal noise in the radar receiver and Rayleigh fading, the signal power measurements are corrupted by noise. Typically, a separate measurement of the receive-only noise power is made and subtracted from the signal-plus-noise measurement to yield a backscatter signal power measurement. The measurement signal-to-noise ratio (SNR) depends on the noise level, radar parameters, and the scene but may vary from very low (< -40 dB) to high (20 dB or more).

Scatterometers are generally used with distributed (areal) targets. For such a target the backscattered signal power P_s is related to the normalized radar cross section (σ^0) via the radar equation (Ulaby et al., 1981)

$$\sigma^0 = P_s P_t G^2 \lambda^2 A_c / (4\pi)^3 R^4 L \quad (1)$$

where P_t is the transmitted power, G is antenna gain, λ is the wavelength of the transmitted microwave signal, A_c

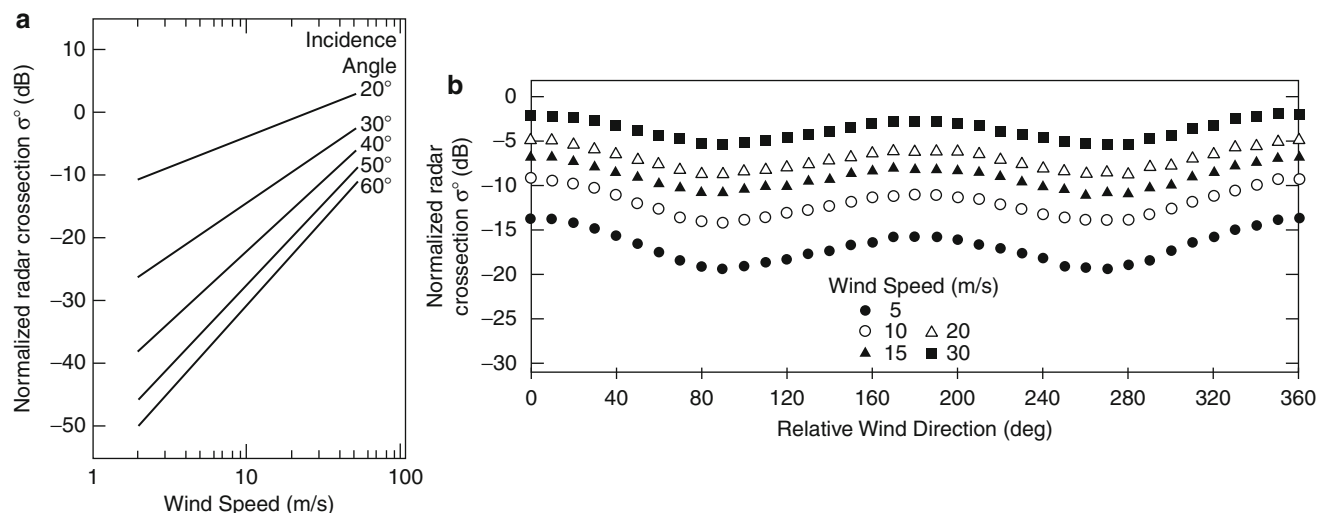
is the resolved target area, R is the slant range to the surface, and L is the net system losses. For a given target, the measured radar cross section depends on the observational geometry including the incidence angle and (possibly) the azimuth angle. The cross section is related to the dielectric properties, surface roughness, and volume scattering characteristics of the target. With careful calibration and careful design to ensure measurement stability, the normalized radar cross section can be very accurately estimated, e.g., to better than ± 0.1 dB. The measurements of the cross section are then used to estimate target characteristics of interest.

Scatterometer transmit signals can be modulated pulses or interrupted continuous wave (ICW) signals. Processing of the received signal can yield cross-sectional measurements at finer scale than the antenna footprint. Pulse compression or matched filtering is used for this when the transmit signal is modulated, while Doppler filtering is used for ICW signals. In either case, signal power measurements from several pulses are averaged into a signal power measurement.

The backscattered signal from the target scene consists of distance-attenuated, time-delayed, and frequency-shifted copies of the transmitted signal. The time delay arises due to the finite speed of light and the distance (sometimes referred to as the slant range) between the radar and the scene. Scatterometers can be designed for stationary operation or for operation from moving platforms such as spacecraft. The relative motion of the radar with respect to the scene introduces Doppler frequency shifts. This relative motion coupled with the coherent nature of radar results in Rayleigh fading (Ulaby et al., 1981) or signal scintillation due to self-interference of signal from different component areas of the target. Rayleigh fading introduces signal variability that is often treated as undesired “noise” in the signal power measurement.

Scatterometer applications

The primary application of radar scatterometers is satellite-based ocean vector wind measurement. Scatterometer-derived winds are operationally used in [weather forecasting](#) and scientific research. Winds blowing over the ocean generate small, centimetric-scale waves known as capillary-gravity waves. These interact nonlinearly to generate larger waves known as swell. For moderate-to-large incidence angles (15° – 60°), the ocean surface backscatter is dominated by Bragg scattering from the capillary-gravity waves. The Bragg scattering results in a strong wind-speed and direction-dependent variation in the radar backscatter which is exploited to estimate the wind from the backscatter measurements. The relationship between the radar backscatter and the wind speed and direction is known as the geophysical model function (GMF). The GMF at 14 GHz is shown in [Figure 1](#). By making multiple measurements of the surface backscatter from multiple



Radar, Scatterometers, Figure 1 A representation of the Ku-band (14 GHz) geophysical model function relating the near-surface wind over the ocean to the ocean's normalized radar cross section. (a) Radar cross section versus wind speed and incidence angle. (b) Radar cross section versus the direction difference between the wind and the radar azimuth angle.

azimuth angles, the near-surface wind speed and direction can be determined by inverting the GMF. This estimation process is often termed wind retrieval.

Unfortunately, the double cosine dependence of the radar cross section on the wind direction gives rise to so-called ambiguities, where two possible wind vectors “explain” the cross-sectional measurements. To resolve this ambiguity, ambiguity selection algorithms (also termed “dealiasing”) are employed. Such algorithms generally start with a first guess of which ambiguity to use and then apply wind field continuity considerations to select a unique wind vector at each observation point.

Wind scatterometers measure the radar cross section from multiple azimuth angles on a grid of wind vector cells (wvcs) that are typically 12.5, 25, or 50 km on a side. Typical coverage swaths are 500–1,600 km wide for spaceborne scatterometers and result in global coverage of most of the Earth's oceans from 1 day up to several days.

Scatterometer cross-sectional measurements are also used for scientific studies of vegetation and polar ice. In the case of vegetation, the radar cross section is dependent on canopy roughness, density, and moisture content. Unlike the ocean radar cross section, which is primarily due to surface scattering, the cross section of vegetation is dominated by volume scattering from individual leaves in the canopy. Volume scattering is also important for ice and snow sensing. In this case, backscattering mostly comes from buried ice crystals and structures. When liquid water is present in the snow, the radar cross section changes dramatically when compared to dry conditions. As a result, the primary scattering mechanism shifts from volume to surface scattering, significantly altering the backscatter dependence on incidence angle.

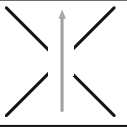
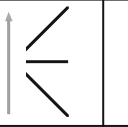
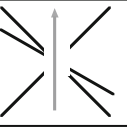
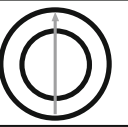
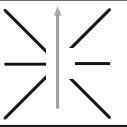
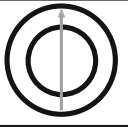
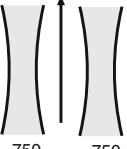
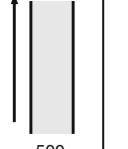
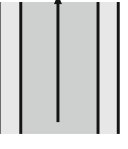
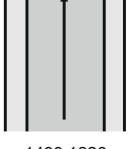
This sensitivity makes scatterometer measurements very useful for monitoring freeze and thaw cycles in the polar regions.

Spaceborne scatterometers

A number of scatterometers have flown onboard Earth-orbiting spacecraft. The first of these was the Seasat-1 Scatterometer (SASS) in 1978, which first demonstrated the utility of spaceborne scatterometers. SASS was followed by the ERS-1 and -2 Advanced Microwave Instrument scatterometer mode (ESCAT) in 1992 and 1995, the NASA scatterometer on ADEOS-I in 1996, the SeaWinds scatterometer on QuikSCAT and ADEOS-II in 1999 and 2003, respectively, and the Advanced Scatterometer (ASCAT) on MetOp A in 2007. SeaWinds on QuikSCAT ceased its primary wind mission in 2009. ASCAT continues operating as of mid-2012. SASS, NSCAT, and SeaWinds were developed by the USA and operated at Ku-band (14 GHz), while ESCAT and ASCAT were developed by the European Space Agency (ESA) and operated at C-band (5.4 GHz). The Indian Space Research Organisation (ISRO) launched a scatterometer similar to QuikSCAT on Oceansat-2 in 2009. The Chinese Space Agency also launched a scatterometer similar to QuikSCAT on HY-2A in 2010.

A summary comparison of the key features of various wind scatterometers is provided in Figure 2.

SASS, ESCAT, NSCAT, and ASCAT were fan-beam scatterometers, while the SeaWinds scatterometer employs a rotating dual pencil-beam antenna. In a fan-beam scatterometer, along-track resolution results from a combination of a narrow antenna pattern and the timing

	SASS	ESCAT	NSCAT	SeaWinds	ASCAT	Oscat
FREQUENCY	14.6 GHz	5.3 GHz	13.995 GHz	13.4 GHz	5.3 GHz	13.5 GHz
ANTENNA AZIMUTHS						
POLARIZATIONS	V-H, V-H	V ONLY	V, V-H, V	V-OUTER/H-INNER	V ONLY	V-OUTER/H-INNER
BEAM RESOLUTION	FIXED DOPPLER	RANGE GATE	VARIABLE DOPPLER	PENCIL-BEAM	RANGE GATE	PENCIL-BEAM
SCIENCE MODES	MANY	SAR, WIND	WIND ONLY	WIND/HI-RES	WIND ONLY	WIND/HI-RES
RESOLUTION (σ°)	nominally 50 km	50 km	25 km	Egg: 25x35 km Slice: 6x25km	25/50 km	Egg: 30x68 km Slice: 6x30 km
SWATH, km	 ~750 ~750	 500	 600 600	 1400,1800	 500 500	 1400,1836
INCIDENCE ANGLES	0° - 70°	18° - 59°	17° - 60°	46° & 54.4°	25° - 65°	49° & 57°
DAILY COVERAGE	VARIABLE	< 41 %	78 %	92 %	65 %	> 90 %
MISSION & DATES	SEASAT: 6/78 -10/78	ERS-1: 92 - 96 ERS-2: 95 -01	ADEOS-I: 8/96 - 6/97	QuikSCAT: 6/99-11/09 ADEOS-II: 1/02-10/02	METOP-A: 6/07- METOP-B: 4/09-	OceanSat-2: 10/09-

Radar, Scatterometers, Figure 2 A comparison of key features of various spaceborne wind scatterometers.

of transmit pulses integrated into a single measurement cell. Cross-track resolution is obtained either by range gate filtering (ESCAT and ASCAT) or by Doppler filtering (SASS and NSCAT) and the narrow beam pattern. Multiple fan-beam antennas enable multiple azimuth observations over a variety of azimuth angles.

The helically scanning SeaWinds antenna makes measurements at fixed incidence angles but at a wide diversity of azimuth angles. Measurement resolution for the SeaWinds scatterometer is based on the pencil-beam antenna footprint. Doppler processing further refines the measurement resolution. The Oceansat-2 and HY-2A scatterometers operate similarly.

Reported backscatter resolution for these sensors varies from 6 km by 25 km (SeaWinds) to 50 km SASS. Backscatter measurements are grouped on to 12.5, 25, or 50 km wind vector cell grids, depending on the sensor, and winds are estimated for each wvc.

Conclusion

Though primarily deployed for ocean wind measurement, scatterometer backscatter measurements have also proven useful in studies of land, vegetation, and polar ice. Spaceborne scatterometers have proven to be very powerful remote sensing instruments with measurements that span multiple decades and long, nearly continuous global data sets of well-calibrated backscatter and derived products. Scatterometer thus provide a valuable data set for studying global change.

Bibliography

- Attema, E., 1991. The active microwave instrument onboard the ERS-1 satellite. *Proceedings of the IEEE*, **79**(6), 791–799.
- Elachi, C., 1988. *Spaceborne Radar Remote Sensing: Applications and Techniques*. New York: IEEE Press.
- Figa-Saldaña, J., Wilson, J. J. W., Attema, E., Gelsthorpe, R., Drinkwater, M. R., and Stofflen, A., 2002. The advanced scatterometer (ASCAT) on the Meteorological Operational (MetOp) Platform: a follow on for European wind scatterometers. *Canadian Journal of Remote Sensing*, **28**(3), 404–412.
- Graham, W. L., et al., 1977. The Seasat-A satellite scatterometer. *IEEE Journal of Oceanic Engineering*, **OE-2**, 200–206.
- Long, D. G., Drinkwater, M. R., Holt, B., Saatchi, S., and Bertoin, C., 2001. Global ice and land climate studies using scatterometer image data. *EOS, Transactions of the American Geophysical Union*, **82**(43), 503.
- Naderi, F., Freilich, M. H., and Long, D. G., 1991. Spaceborne radar measurement of wind velocity over the ocean – an overview of the NSCAT scatterometer system. *Proceedings of the IEEE*, **79**(6), 850–866.
- Spencer, M. W., Wu, C., and Long, D. G., 2000. Improved resolution backscatter measurements with the seawinds pencil-beam scatterometer. *IEEE Transactions on Geoscience and Remote Sensing*, **38**(1), 89–104.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing – Active and Passive, Vols. 1 and 2*. Reading, MA: Addison-Wesley.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1990. *Microwave Remote Sensing – Active and Passive*. Reading, MA: Addison-Wesley, Vol. 3.

Cross-references

[Calibration, Scatterometers](#)

RADAR, SYNTHETIC APERTURE

Keith Raney

Applied Physics Laboratory, Johns Hopkins University,
Laurel, MD, USA

Synonyms

Active sensor; DInSAR; InSAR; Radar interferometry;
Radar polarimetry; Range radar; SAR

Definition

A synthetic aperture radar (SAR) consists of two logical functions: radar and processor. The *radar* is an active imaging sensor that looks to the side as it moves along its trajectory and collects reflections from transmissions pulse by pulse, storing these data in memory. An individual reflecting object contributes reflections so long as it remains illuminated by the radar's antenna pattern. The *processor* coherently integrates these data using an algorithm that emulates the action of an antenna whose along-track aperture length equals the length of the data record from an individual backscatterer. The data record encompasses the signals that would have been received by a very long real antenna aperture, which is the source of the term "synthetic (or synthesized) aperture." The along-track (or, equivalently, the cross-range or azimuth) resolution that results is comparable to half of the along-track dimension of the radar's actual antenna. The radar must be *coherent* – thus phase preserving – while it collects reflections, so that the phase structure of the signal ensemble in memory from each and every scatterer is equivalent to that of a wavefront that would result from a single transmission.

Introduction

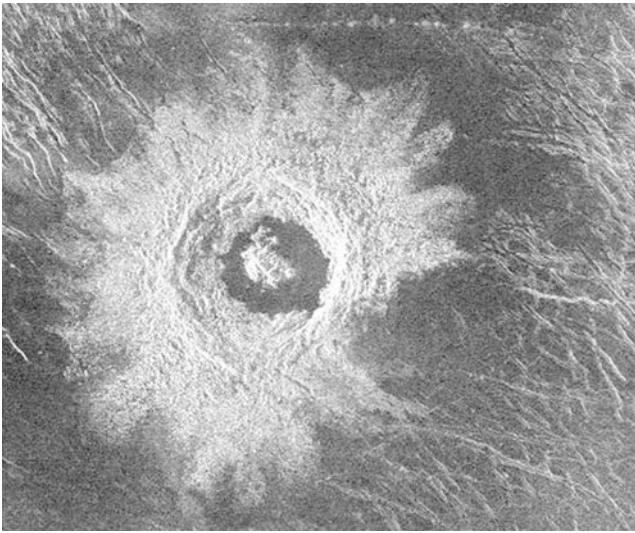
In its original and simplest form, a synthetic aperture radar (SAR) is a radar whose output product is an image – a two-dimensional mapping of "radar brightness" that resembles a black-and-white aerial photograph. The signal's "phase history" is the essential input to the processor that generates SAR images. More elegant forms of SAR systems take further advantage of the phase structure embedded in the data records to extract quantitative geophysical information about the scene. This entry is designed to provide a glimpse into these basic capabilities of SAR systems and to highlight key aspects of the underlying physics. [Table 1](#) provides a brief list of pertinent historical milestones relevant to modern SAR. For more information on specific SAR systems, the reader is advised to follow the citations in the discussion and to consult the (reasonably) complete list in Raney (2008). The entry closes with suggestions for tutorials, treatises, and other tools specific to remote sensing SAR systems that may be of interest to the reader.

Radar, Synthetic Aperture, Table 1 Selected milestones (citations in the text) from the history of SAR, including early discoveries that helped to establish SAR's foundations

Year	Source	Accomplishment
1842	Christian Doppler	Effect of relative motion on light
1873	James Clerk Maxwell	Treatise on electricity and magnetism
1888	Rayleigh	Concept of aperture diffraction limit
1888	Heinrich Hertz	Demonstrated radio waves; measured "c"
1904	Christian Hulsmeyer	Patented "radar"; 1 mile demonstration
1934	Sir R. Watson-Watt	British chain home
1951	Carl Wiley	Resolution $\sim 1/\text{bandwidth}$ $\sim$ "synthetic aperture"
1953	U. Illinois	First strip "SAR" image
1955	Emmett Leith	Doppler linear fm $\sim$ Fresnel lens
1957	U. Mich (WRL)	First fully focused strip SAR = > AN/UPD-1
1958	Willow Run Lab	Range chirp (after Darlington) $\sim$ Fresnel lens
1962	Emmett Leith	"Free space $\sim$ pulse compression filter"
1963	Dwight North	Matched filter (after a classified 1943 report)
1971	William Caputi	"Stretch" long linear deramp
1974	Jack Walker	SpotSAR (earlier Ph. D.; first open publication)
1978	NASA	Seasat (SAR + other instruments)
1978	MDA and Canada (CCRS)	First digital processor for Seasat SAR
1981	J. P. Claasen	ScanSAR idea (student of R. K. Moore, U. Kans)
1983	USSR	Venera radars at Venus
1985	USA/JPL	SAR polarimetry; 2-pass interferometry
1990	USA	Magellan at Venus
1991	Europe	ERS-1: first operational (civilian) space SAR
1995	Canada	First multimode space SAR (RADARSAT)
1996	Cloude and Pottier	Polarimetric decomposition
2000	USA/Germany/Italy	Shuttle radar terrain mapper (topography)
2006	Japan	First polarimetric space SAR (ALOS/PALSAR)
2008	USA/India	First hybrid-polarity (lunar) SAR

Imagery

Magellan – the 1990 NASA mission to Venus (Johnson, 1991; Pettengill et al., 1991) – successfully mapped more than 98 % of that planet's cloud-shrouded surface at nominally 120 m resolution (75 m pixels). An example is shown in [Figure 1](#). In such imagery, white represents radar-bright backscatter, as opposed to darker (black) areas that have less radar backscatter. Over terrain such as this, radar-bright areas correspond to surface roughness whose scale is on the order of the wavelength, which for Magellan was about 12 cm (S-band) (see "[Radars](#)").



Radar, Synthetic Aperture, Figure 1 Magellan craters. The impact crater Golubkina on the surface of Venus imaged by the Magellan S-band, HH-polarized SAR. Rough ejecta give rise to strong radar return, a blessing for those interested in planetary geology (Image credit, NASA).

SAR imagery appears to be viewed from above, even though the radar illuminates the scene from the side. This perspective is markedly different from that of conventional optical sensors, virtually all of which may be classed as “angle-angle” imagers. In both dimensions of an optical system, there is a ray-trace relationship between objects in the field of view and their portrayal in the corresponding focal plane. In contrast, a SAR is an example of a range-Doppler imager. The along-range (“line-of-sight”) coordinate system is distance, proportional to the round-trip time delay of the illuminating pulse and the resulting reflections. The cross-range dimension of a (side-looking) SAR is set up by the sensor’s along-track velocity. The resulting imaging perspective (Figure 2) is equivalent to that of an observer located at right angles to the slant-range plane, which is the plane that contains the radar’s radiation toward the scene. Several consequences follow from this perspective, including elevation displacement and layover leading to apparent distortions of taller features in SAR imagery and the appearance of radar shadows (Henderson and Lewis, 1998).

Phase: the fundamental measurement

An optical imager such as a common camera “works” because its lens imposes a phase delay (proportional to its thickness) that converts the spherical divergence of the incoming EM wave field into a corresponding convergence of the outgoing wave field. The camera’s film (or CCD array) is placed at the focal point of that convergence. The lens’ shape is dictated by the desired phase delay – thicker in the middle where the incoming wave

components need to be retarded and thinner toward the perimeter where the wave components need less delay. Data redistribution in the focal plane is the Fourier transform of the data across the wave emerging from the lens (Born and Wolf, 1959; O’Neill, 1963). The operations of a SAR processor are the same in principle. The processor imposes phases onto wavefront constituents in memory according to their relative position along the synthetic aperture (which may be visualized as a virtual microwave lens) and then collects (by integral transformation) the result onto a virtual focal plane.

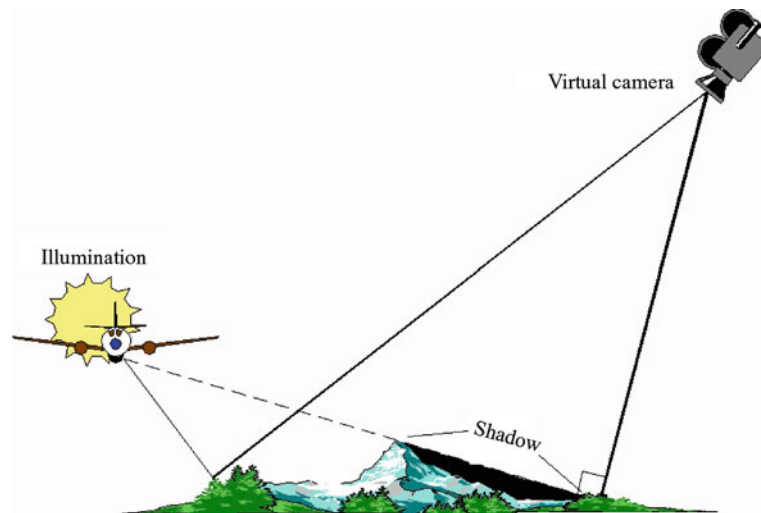
Figure 3 shows the wavefront phase history of an individual reflecting object. The controlling system parameters are radar velocity V and (minimum) range R , radar wavelength λ , and the along-track location x of an individual scatterer with respect to the radar’s minimum range. At all scatterer positions, the incremental range $\Delta R = x^2/2R$ (following a small angle simplification, which often is reasonable for SAR systems). In radar, phase and range are related by the proportionality $2k$, where $k = 2\pi/\lambda$ is the usual wave number. Then

$$\text{Phase_increment} = 2k\Delta R \quad (1)$$

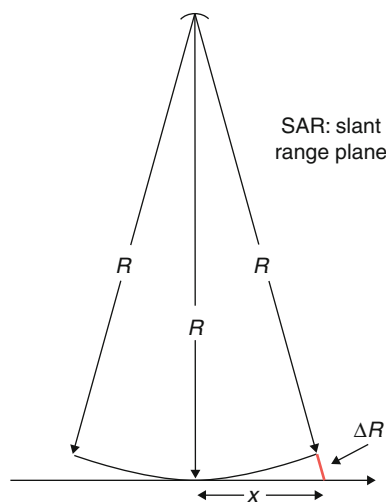
which could be defended easily as the most fundamental equation in the theory of SAR! (If ΔR is evaluated using $x = Vt$, then an individual scatterer’s phase variation as it moves across the radar’s field of view is seen to be quadratic, for which its time derivative is linear in time, which is the traditional frequency characteristic of SAR.) Note that it is not necessary that the ΔR ensemble for a given scatterer corresponds to a spherical wavefront nor that samples of ΔR be uniformly distributed along the trajectory of the radar. All that is required is that the processor “know” the incremental phases and where they are located. This degree of freedom enables effective processing of SAR data from radars mounted on maneuvering aircraft for example.

Processor: the essential ingredient

A SAR processor has essentially two tasks: image focus and data product generation. Image focus always requires azimuth processing and usually requires range processing as well. SARs use some form of range pulse coding – often linear frequency modulation (FM) – to increase the effective energy transmitted, thus improving the system SNR (see “Radars”). Linear FM is one-to-one equivalent to the linear FM of the azimuth dimension of SAR data and so may be processed in similar fashion. This similarity was first pointed out by Emmett Leith in his brilliant conceptualization of optical processing for the first successful airborne SAR system (Cutrona et al., 1960; Leith, 2000; Leith and Upatnieks, 1965). However, other range modulation schemes may be used, such as the binary code of Magellan’s radar. No matter the range modulation method, however, the SAR processor must include range focus (unless the radar includes range compression, as is sometimes true). Range processing often is based on the



Radar, Synthetic Aperture, Figure 2 Image viewing perspective. The range coordinate independent variable for the image from a side-looking radar is proportional to round-trip time delay; hence, the viewing perspective appears to be at a right angle to the slant-range plane.



Radar, Synthetic Aperture, Figure 3 Range increment. As the SAR passes a given object in the scene, at every angle either side of minimum range, the range is larger by a small increment. The SAR measures the relative phase of this extra distance. The resulting phase history is the prime input to the SAR image processor.

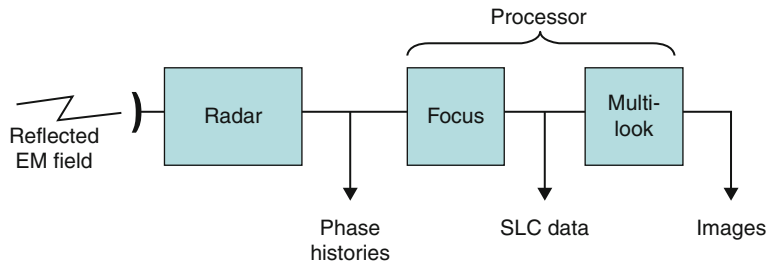
matched-filter paradigm (North, 1963; Turin, 1960), according to which the data are Fourier transformed, multiplied by a conjugate phase operator, and then inverse Fourier transformed.

Azimuth processing must be matched to the phase histories of the data. This simple statement implies several significant consequences. In any practical setting, there are many scatterers at different range and along-track positions simultaneously illuminated by the radar. Hence, the all-important range phase increments associated with each

scatterer overlap, so they cannot be phase matched simultaneously in the time domain. One effective strategy to overcome this challenge is to perform along-track Fourier transforms, thus remapping the azimuth data of each such block of data into the corresponding Doppler domain. In this domain, the range phase increments are single valued in azimuth frequency, since all data have their “zero” positions aligned, relative to “zero Doppler.” In principle, the matched-filter approach applies again, noting that starting with data already in the Doppler domain, only the conjugate phase multiply and the inverse Fourier transform are required.

The processor must accommodate the fact that the Doppler FM rate of each scatterer depends on its range R . Further, if the range increment size exceeds the radar’s range resolution, then the processor must correct the resulting range curvature.

The Seasat SAR (Jordan, 1980) is a good example to illustrate several key points. At a range of 900 km, Seasat’s 10 m-long SAR antenna pattern width was about 20 km, sampled by the radar’s pulse repetition frequency (PRF) at spatial separations along track of about 5 m. It follows that there were 4,000 azimuth lines in each synthetic aperture data set, thus 4,000 individual phase increments to match. Seasat’s single-look azimuth resolution was about 5 m (half of the antenna length), thus corresponding to an azimuth compression ratio (see “*Radars*”) of 4,000. At the beam edge, where the scatterers would be located 10 km from their minimum range, the corresponding range increment was 1/18 km, or about 55 m, which was on the order of 4 times larger than the radar’s (slant-range) resolution. Consequently, processing Seasat’s SAR data required range curvature correction (typically implemented by integral shifts and phase multipliers).



Radar, Synthetic Aperture, Figure 4 High-level block diagram. A SAR consists of two parts: the radar and the processor. The processor focuses the data collected by the radar, known as the single-look complex (SLC) image. The processor may also detect the SLC (estimate its power pixel by pixel), then form averaged (multi-look) imagery, which is the usual visual product.

By convention, the result of two-dimensional focusing is denoted as single-look complex (SLC) data, which consist of digital numbers that represent the image's amplitude and phase at each picture element location (pixel). In principle, these data are a perfect linear transformation of the EM wave field excited by the radar's transmissions. SLC data is most useful for subsequent processing (e.g., interferometry) that depends on signal phase information. The second task of a SAR processor is to generate the data products, which usually are one of two types, as suggested in Figure 4. Multi-look imagery is the more familiar form of SAR data, such as the example from Magellan in Figure 1. Imagery is generated within the processor by (1) detection (usually magnitude squared) to transform the SLC data into radar brightness, thus destroying the phase information, and (2) multi-look summation. The radar brightness of SLC data has very poor precision (see "*Radars*"), since by definition, it is characterized by a mean-squared-to-standard-deviation ratio of unity. This is due to the coherent nature of the input data, which generates manifold self-interference within the resulting imagery; the resulting "noise" is known as SAR speckle. Speckle has its good aspects as well as bad (Raney, 1998). General practice is to reduce speckle either by averaging statistically independent versions of the same scene or by averaging adjacent areas of a detected SLC image. Typical standard SAR image products have four or more looks.

For many years, there has been interest in filters that suppress speckle. If linear techniques are used, then there always must be a trade-off between looks and resolution (Porcello et al., 1976). Nonlinear techniques promise to get around that fundamental constraint, but the results are mixed. Early versions imposed a bias on the local mean and had difficulty with bright point reflections. Lee et al. (2008) have developed a mean-preserving technique that manages bright targets well and which may be implemented with an efficient algorithm.

For more on SAR processing, see Curlander and McDonough (1991) and Data Processing, SAR Sensors.

SAR interferometry

A SAR image of terrain is a two-dimensional mapping of a three-dimensional surface. Unless more information is

available, there is no way to quantitatively estimate terrain height with respect to a datum plane. Of course, as with aerial photographs, a pair of SAR images, each taken at different incident angles, could be used as a stereo pair. Being a coherent system, however, SAR supports interferometry, which is a much more powerful tool.

The basic interferometric environment is one in which there is a pair of mutually coherent data ensembles having the same carrier frequency and two-dimensional phase modulations, but within which a subset of elements have different reference phases (Hanssen, 2001). Signal processing is designed to estimate those phase differences and to deduce geophysical parameters of interest from the resulting measurement. In its finest forms, SAR interferometry is able to generate high-quality elevation contour maps or to observe physical movements or other changes in the observed scene that are at scales smaller than the wavelength of the radar being used, regardless of the radar's resolution.

The interferometric signal model is simple in concept. For the SLC images from one scattering element, let the signal pair be described by:

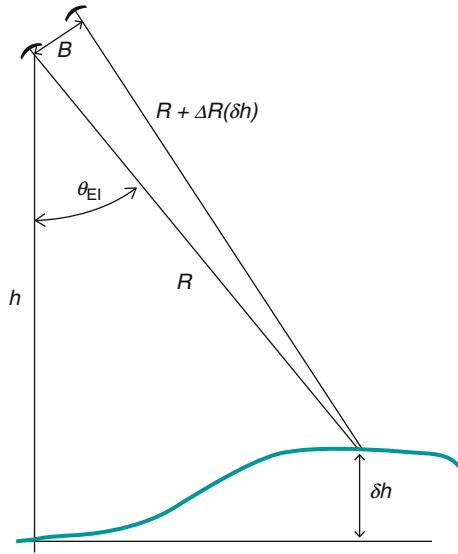
$$\begin{aligned} s_1(t) &= \Gamma_1 \exp\{-j\varphi\} \text{ ("master")} \\ s_2(t) &= \Gamma_2 \exp\{-j\phi + j\Delta\varphi(t_2 - t_1)\} \text{ ("slave")} \end{aligned} \quad (2)$$

where the subscripts t_1 and t_2 suggest that the two signals may be obtained at two different points in time as well as from two different spatial perspectives. The objective is to estimate (usually by cross-correlation) the relative interferometric phase difference:

$$\text{Phase_difference} = \Delta\varphi(t_2 - t_1) \quad (3)$$

The phase difference may be due either to geometric or to temporal differences (at wavelength scales) between the two observations.

Any approach to phase difference measurement is subject to the fundamental 2π ambiguity characteristic of phase estimation algorithms. In many radar applications, knowledge of the physical constraints of the situation coupled with phase unwrapping algorithms is sufficient for the purpose. More on the phase unwrapping problem may be found in Zebker and Lu (1998).



Radar, Synthetic Aperture, Figure 5 Elevation baseline. Data gathered by a radar from two vantage points separated in the vertical plane are sufficient to estimate the range elevation angle as well as the range distance to each resolved scatterer. The sensitivity of this vector range measurement depends on the radar's wavelength, baseline B , and viewing angle θ_{EI} .

Successful interferometry requires at least some mutual coherence between the complex scattering coefficients Γ_1 and Γ_2 . Coherence usually is expressed in the classic form (Born and Wolf, 1959) as the normalized mutual coherence function γ_{12} . A map of coherence over a scene of interest frequently is a supporting data product for interferometric SAR, where $\gamma_{12} \sim 1$ indicates high coherence, therefore more robust interferometry, and smaller γ_{12} indicates less coherence.

Topographic contours

Consider two SAR images of the same scene taken by the same radar from the same aspect, but from different altitudes, as shown in Figure 5. The slight difference in range between the two views differs across the scene on the scale of the radar's wavelength, a difference that is expressed in terms of phase and, in particular, in the phase exposed by interferometric combination of the two (SLC) data sets. The phase distribution over the scene is known as an interferogram, which can be interpreted as a mapping of the elevation contours of the surface, once the systematic phase grid due to the elevated viewing geometry is removed, and the 2π (phase-wrap) ambiguities are sorted out. The first demonstration of this technique was from an aircraft, with two antennas, one mounted above the other (Graham, 1974). That antenna separation of a meter or so may be fine for an aircraft, but the optimal vertical separation is proportional to radar range, leading to a desired interferometric baseline of 50 m or more for an orbital system, which for most spacecraft would not be practicable. The Shuttle Radar Topography Mission (SRTM) provided a very nice

example of space-based single pass SAR interferometry (Bamler et al., 2003; SRTM, 2005).

In the mid-1980s, a brilliant innovation was proposed by R. Goldstein (Zebker and Goldstein, 1986) to circumvent the antenna separation problem for satellite-based radars, namely, to use the same radar but on different orbital passes to get the required master/slave data pair, under the constraint that the two data sets must be mutually coherent. This technique, known as repeat-pass InSAR, has become the mainstay of space-based SAR operations.

The constraint on mutual coherence is not trivial, with spatial and temporal implications. Assume at the outset that the scene reflectivity is the same for both passes of the radar, particularly at the detailed level of the radar wavelength. The spatial constraint applies to the effective baseline B between the two orbital passes. For full coherence, the radar wavelength projected onto each area of the surface must be the same from both orbits. Since the two orbits are separated, the surface is observed at a slightly different incident angle, so that the wavelengths projected onto the surface are proportionally different. Mutual coherence is supported only if the range bandwidth of the radar signal is sufficient to assure overlap of the spectra of the wave numbers as projected onto the surface. Loss of mutual coherence through increasing orbit spacing is known as *baseline decorrelation* (Gabriel and Goldstein, 1988; Zebker and Villasenor, 1992). One may show for reasonably level terrain that the upper bound on acceptable baseline is:

$$B \leq R\lambda \tan \theta_{EI} / 2r_R \quad (4)$$

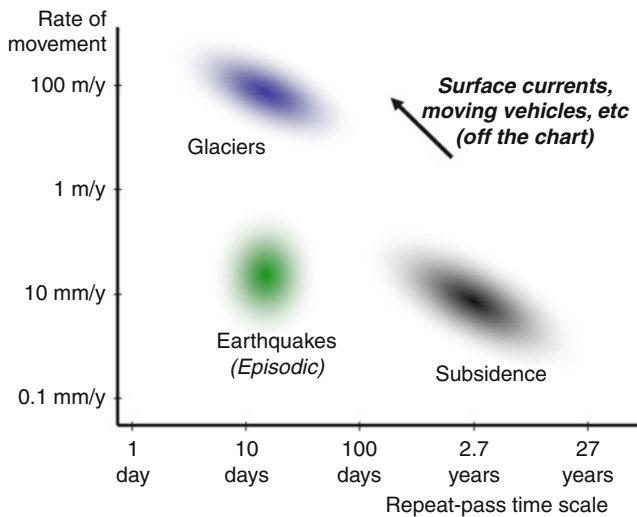
where r_R is slant-range resolution (which is inversely proportional to range bandwidth) and θ_{EI} is the angle between the radar line of sight and vertical (toward nadir from the spacecraft).

The second consideration has to do with scene coherence. This requirement is more readily satisfied for short inter-opportunity intervals and for longer-wavelength radars. For many scenes, one can find the so-called permanent (or persistent) scatterers (Ferretti et al., 2001), features (such as corner reflecting shapes that are frequently seen in urban areas) that maintain their phase characteristics seemingly indefinitely.

Coherent movement detection

Now consider two images of the same scene taken at two different times, but from (exactly) the same vantage point. If there is movement of even one coherent scatterer within the scene between the two observations, this may be evidenced as a change in the reference phase of the corresponding signal and hence detectable. An element of the scene with velocity v_{rad} toward the radar over an observation time interval of δt seconds will result in:

$$\text{Movement_phase_difference} = 2kv_{rad}\delta t \quad (5)$$



Radar, Synthetic Aperture, Figure 6 Motion sensitivity. The sensitivity of multi-pass SAR data sets to coherently sensible motions in the observed scene varies over several orders of magnitude, as illustrated by the examples of glaciers, earthquakes, and land subsidence.

which becomes more sensitive to smaller velocities for shorter wavelength (recall $k = 2\pi/\lambda$) and longer observation time. This technique lies behind all reasonable SAR-based moving-target-indicator radars (Raney, 1971; Urkowitz, 1964) and has been demonstrated from space-based systems (Palubinskas and Runge, 2008).

There is a huge range of possibilities lurking in Equation 5, from relatively high-speed movements such as vehicles or ocean currents (which require observation time intervals as short as 1 ms) to glaciers (Venkataraman et al., 2005) or even land subsidence (URL/PSInSAR, 2008) (Figure 6). The resulting data products are maps that show the distribution and rates of movement of elements within the scene. The same technique also can display in graphic detail differential land movements caused by an earthquake (Massonnet et al., 1993), if a pair of suitable SAR data are available from passes before and after the event.

In practice, it usually is not possible to repeat the same measurement from exactly the same vantage point, so that there will be residual topographic signals in the interferogram along with coherently sensible displacement signals. The topographic component can be estimated and removed by using a third image in combination with one of the other two, creating a pair for which the temporal baseline is too short to be sensitive to slow small in-scene movement, yet whose spatial baseline is suitable for terrain elevation measurement.

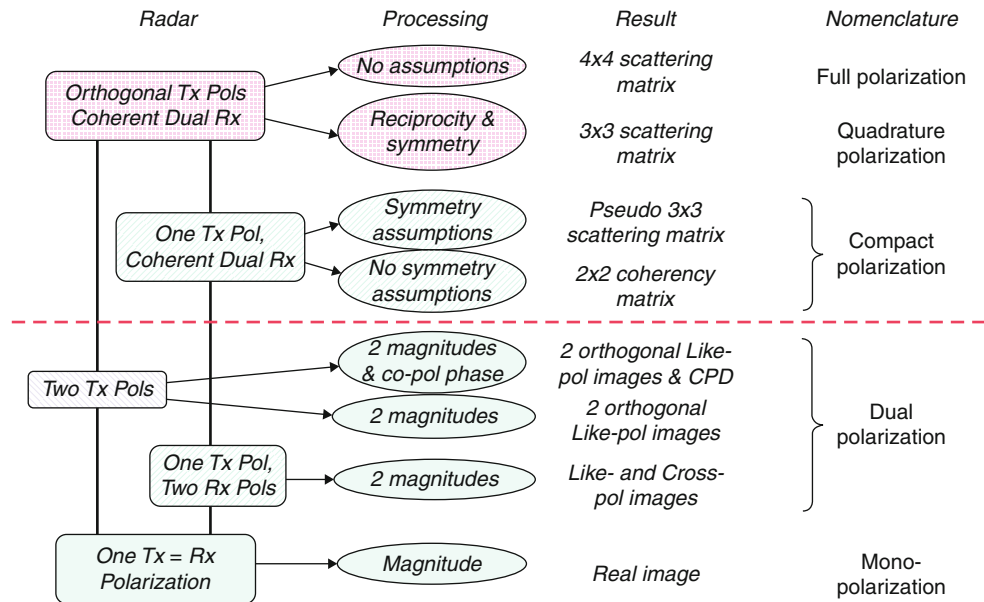
SAR polarimetry

A polarimetric radar is one that measures certain parameters of the observed EM field that depend on the transmitted and/or received polarizations (see “*Radiation, Polarization, and Coherence*”). The first dual-polarized imaging radar

was the APQ-97, which offered photographic images at either HH or HV polarization. Since the 1950s, the variety of transmit and receive polarization combinations offered by SAR systems has expanded considerably, for which a road map (Figure 7) should be helpful. Single polarization is the simplest case, epitomized by the ERS-1/2 (Attema, 1991), Magellan (Johnson, 1991), and RADARSAT-1 (Raney et al., 1991) systems. Dual polarization usually refers to two receive polarizations following one transmitted polarization, in the same spirit as the APQ-97; hence, the relative phase between the received polarizations is not available. In this road map, “dual polarization” also includes two transmitted polarizations when they alternate noncoherently, illustrated by ENVISAT (URL/ASAR, 2008) and certain modes on ALOS/PALSAR (URL/PALSAR, 2008). The distinguishing feature of this class of polarimetric radars is that their primary data product is *imagery*, sufficient to calculate ratios of powers drawn from different polarization combinations. Conventional dual-polarized imaging radars do not measure relative phase between the two receiver channels.

As was shown by Stokes in 1852 (Stokes, 1852), any partially coherent quasi-monochromatic EM field can be fully characterized by four real numbers, known as the Stokes parameters. These may be derived from the 2×2 coherency matrix of the field or, equivalently, from its covariance matrix. This theme characterizes the polarization combinations of Figure 7 whose data products may be viewed as matrices. Evaluation of the elements in these matrices requires knowledge of the relative phase between the radar’s two receive channels. This class of radars makes polarimetric measurements, not just polarization-dependent images.

The most complete embodiment is fully polarimetric, in which the transmitted field alternates between two orthogonal polarizations at Nyquist rate (which assures that the Doppler bandwidth of the backscatter is sampled adequately by both polarizations) and the resulting backscatter is received on two orthogonal polarizations coherently [preserving the relative phase between the two channels]. The classic form (van Zyl et al., 1987) transmits H or V polarization, multiplexed, and receives H and V coherently after each transmission. It turns out that a hybrid-polarization architecture has several advantages over linearly polarized architecture (Freeman and Raney, 2008; Raney, 2007). Hybrid polarization denotes an architecture in which circular polarizations are transmitted (L or R) and linear polarizations are received (H and V). Independent of the polarization plan, the resulting data from a fully polarimetric radar are sufficient to calculate the 4×4 covariance matrix representation of the data, which in turn is sufficient to fully characterize the complex reflectivity coefficient of all resolved elements in the scene. Certain symmetries may be found in the 4×4 matrix, which when exploited under appropriate assumptions lead to a reduced 3×3 matrix, commonly accepted as the standard “quad-pol” data product (Zebker and Van Zyl, 1991).



Radar, Synthetic Aperture, Figure 7 Hierarchy of polarimetric SARs. There are many possible combinations of receive and transmit polarizations and associated terminology for SAR systems. This chart traces the relationship between the radar and its data products for a single-polarization system, and progressively more elaborate SARs, from dual polarization through compact to full polarization.

If only one polarization is transmitted and two polarizations are received coherently, then the natural data product is the 2×2 covariance matrix of the observed field. (The elements of the complex scattering matrix cannot be found, because the single transmit polarization is not sufficient to probe all possible transmit and receive reflectivities.) This type of coherently dual-polarized radar is known as compact polarimetry, since its objective is to enjoy many of the benefits of a quad-pol system without incurring its disadvantages (Dubois-Fernandez et al., 2007; Souyris and Mingot, 2002). The 2×2 covariance matrix can be expanded to a 3×3 pseudo-covariance matrix by postulating certain symmetry properties (Souyris et al., 2005). The trick with a compact-pol radar is to choose the transmit polarization to match the measurement objectives of the radar; two appealing options are $\pi/4$ (Souyris and Mingot, 2002), known as slant polarization in radar meteorology, and circular (Raney, 2007).

Elegant analysis tools have been developed for quad-pol data (Cloude and Pottier, 1996; 1997), based in large part on well-established matrix decomposition techniques borrowed from quantum mechanics. Similar methodology also may be applied to data from the newer compact-pol architectures. These techniques are successful in parsing the scattering characteristics according to several disjoint classes, including volume backscatter, single bounce (including Bragg reflection from a random surface such as the ocean), and double bounce (from certain geological

formations or volumetric ice deposits), all of which have value in specific applications. For an excellent online tutorial, see URL/Polarimetry_Tutorial (2008).

SAR design rules of the road

SAR systems must obey certain physical rules if they are to provide useful remote sensing data. The key issues are summarized in this section. For more on the theory of SAR design, see Curlander and McDonough (1991) and Harger (1974). More radar-specific issues may be found in Skolnik (2008).

PRF constraints

SAR is a sampled system; hence, the sampling – the pulse repetition frequency (PRF) – must satisfy the Nyquist rule. It follows that the PRF f_p be sufficiently high to sample unambiguously the Doppler spectrum of width B_{Dop} . A radar must receive as well as transmit. Thus, the PRF must also be low enough so that there is time between transmissions to receive the data backscattered from the intended swath of slant-range (time-domain) width T_R . The traditional form (Curlander and McDonough, 1991) for these bounds on SAR PRF is:

$$B_{Dop} < f_p < 1/T_R \quad (6)$$

In practice, sufficient margin must be included in both the upper and lower limits to account for the length of

the transmitted pulse and for the fact that neither the spectrum nor the antenna's elevation pattern has sharp cutoffs. Equation 6 implies an intimate relationship between swath width (in range) and resolution (in azimuth), with consequences of course on the top-level characteristic output data products.

The lower bound often is recast as:

$$B_{Dop} = \frac{2V}{D_{Az}} < f_p \quad (7)$$

which states that the PRF must be sufficiently high such that there are at least two transmissions per antenna aperture length D_{Az} as the radar moves along its trajectory at velocity V . Usually the PRF lower bound is set so that there is a margin of 25 % or more with respect to this constraint.

In an airborne SAR, the PRF constraint is derived to satisfy the bandwidth-limited lower bound, from which follows the maximum range that the radar can operate without introducing ambiguities. However, by default, the minimum range for a space-based SAR is its orbital altitude, usually 600 km or higher. The typical slant range to the intended scene may be 800 km and more. Thus, the upper bound on the PRF should not be set by the range to the scene, but rather by the range width of the area to be imaged. As a consequence, the resulting high PRF will generate a sequence of pulses at any moment that are distributed between the radar and the scene. The space between pulses must be larger than the intended swath width. For example, in certain modes, RADARSAT-2 (URL/RADARSAT, 2008) generates seven pulses "in flight" simultaneously. At the beginning of such a data collection, backscatter from the intended scene would arrive only after the seventh pulse had been transmitted.

Ambiguities

The PRF generates a two-dimensionally sampled space when the data are decomposed into "slow time" (in the azimuth direction) and "fast time" (in the range direction). In azimuth, the PRF creates aliased versions of the data illuminated by the main beam of the antenna. The spectra of these aliases are located at multiples of the PRF to either side of the centroid of the main beam. Of course, when sampled, they are folded back into the Nyquist passband. These aliases are azimuth ambiguities, which should be suppressed – hence not visible – in a well-designed system. This burden falls upon the antenna illumination pattern and frequency-dependent weighting in the processor, as well as on the choice of PRF. Azimuth ambiguities are relatively easy to identify because they are weaker (ghost) duplicates of image features that were collected through the main beam and, therefore, at earlier or later positions along the image strip. The azimuth shift of the ambiguities relative to the central image is an integral multiple of $\Delta X = R\lambda f_p/(2V)$, which is the spatial

offset corresponding to the PRF. Azimuth ambiguities, especially from point targets, have the same FM rate as scatterers within the main beam, and so their focus is preserved through the processor.

In the range direction, one of the consequences of having many pulses in flight at once is that there are echoes from several different ranges that arrive back at the radar at the same relative delay within the range gate as the reflections from the intended swath. If these extra echoes are sufficiently strong, the resulting image artifacts are range ambiguities. Their relative strength depends on the range side-lobe levels of the antenna pattern. Range ambiguities are not as easy to identify as azimuth ambiguities, because they arise from ranges outside of the nominal swath, hence not echoing features that are imaged elsewhere. Range ambiguities by definition arise from ranges that are different from those for which the processor is set, so that range-ambiguous point targets tend to be defocused.

Antenna area constraint

The principal means of suppressing ambiguities is to confine the main beam of the antenna so that the potential sources of azimuth or range ambiguities are not illuminated or at least are illuminated only very weakly through the antenna's side lobes. This requirement imposes a minimum area constraint on the SAR's antenna. The lower and upper bound PRF constraints of B_{Dop} and $1/T_R$ lead to:

$$D_{El}D_{Az} > 4RV(\lambda/c) \tan \theta \quad (8)$$

where the antenna area is the product of its length D_{Az} and height D_{El} , and θ is the mean incident angle in the imaged swath. The range-velocity product in this expression is determined by the parameters peculiar to the particular planet (or moon) about which the SAR is to operate (see Table 2 in "Radars").

Image quality

Engineering norms and geophysical information capacity are the two dominant aspects of image quality for data from remote sensing SARs. Engineering norms include the radar's additive signal-to-noise ratio (SNR) and its multiplicative signal-to-noise ratio (MNR) (see "Radars"). Good SNR means that areas of relatively weak backscatter appear above the radar's intrinsic noise level. Good MNR means that the data products are free of artifacts such as range or azimuth ambiguities.

Geophysical information capacity is equally important as an image quality indicator for remote sensing SAR systems, although it is a more subtle concept. Looks and resolution work together to determine this aspect of image quality, one that is particularly appropriate for SAR response to natural terrain composed of distributed

scatterers. The governing expression is the SAR image quality parameter (Raney, 1998):

$$Q_{SAR} = \frac{N_L}{r_{Rg} r_{Az}} \quad (9)$$

where N_L is the number of (statistically independent) looks and r_{Rg} and r_{Az} are the range and azimuth resolution, respectively, on the surface. The important lesson of this relationship is that a modest increase in number of looks can be applied to offset a corresponding degradation in range resolution. Both looks and resolution require support in bandwidth. It follows that Q_{SAR} is proportional to the product of the range and azimuth bandwidths and hence proportional to the (two-dimensional) information capacity of the radar in the Shannon sense (Shannon, 1948). This principle was applied with great profit to the Magellan SAR design. The image quality of Magellan data varied by no more than about $\pm 2\%$ pole to pole (Johnson, 1991), in spite of large variations in radar range, incident angle, and ground range resolution over that radar's elliptical orbit.

SpotSAR and burst modes

The baseline single-look SAR azimuth resolution is proportional to one over the bandwidth generated by the azimuth beamwidth of the side-looking antenna. The corresponding length of the synthetic aperture is equivalent to the along-track spread of the antenna pattern, which, of course, is proportional to range. Azimuth resolution may be sharpened only by increasing bandwidth, which can be done in one of two ways: increasing the antenna's beamwidth or increasing the spread of aspect angles within which the antenna illuminates a given portion of the scene. The latter is the basis for Spotlight SAR (Jakowatz et al., 1996; Walker, 1980), in which the antenna is steered to dwell on the intended area as the radar passes, thus creating a wider total bandwidth (and a longer synthetic aperture). The trade-off is that adjacent areas along track may not be imaged at all. Increasing azimuth resolution by broadening the antenna pattern, by either reducing the aperture length or spoiling the beam, has several disadvantages, especially reduced antenna gain, thus degrading SNR (see "Radars").

Going the other direction – smaller bandwidth – leads to more coarse azimuth resolution. The bandwidth of the original signal history from a given backscatterer may be reduced by the simple expedient of generating a shorter synthetic aperture than in the usual strip mode, for which the radar is "on" all of the time. This logic leads to the "burst mode," which figures prominently in two forms in space-based SARs. Burst mode means that the radar is operated only during limited periods, rather than continuously. Along a single imaged swath, this implies a reduced data rate, which may be necessary to meet the stringent data-rate requirements confronting planetary or lunar missions. Alternatively, the intervals between bursts may be used to illuminate several different range swaths, thus

expanding the area that may be imaged unambiguously. This is the principle behind the ScanSAR mode (Luscombe, 1988; Moore et al., 1981).

Ambiguity space trade-offs

The fundamental rule is that the image space (illuminated by the antenna) must be "underspread" if ambiguities are to be avoided. The underspread condition (Green, 1968) is that:

$$T_R B_{Dop} < 1 \quad (10)$$

where T_R is the (slant) range swath depth of the antenna pattern and B_{Dop} is the corresponding bandwidth. This constraint (which may be recognized as a rearrangement of the bounds in Equation 6) may be expressed in terms of the antenna's azimuth beamwidth β , the number of azimuth looks N_L , and the duration T_{Az} of illumination time (Raney, 2008) to read:

$$\frac{T_R \beta R N_L}{T_{Az} r_{Az}} < 1 \quad (11)$$

which shows how resolution and target illumination time may be traded against each other while still respecting the fundamental ambiguity avoidance constraints. The formal justification for SAR imaging modes – strip map, SpotSAR, burst, and ScanSAR – follows from this simple inequality.

Fundamental principles

Under certain reasonable assumptions, a SAR obeys several high-level principles (Raney, 1998) that prove to be helpful in analysis and offer insight into how the system works. The assumptions are as follows: (1) The radar uses only one transmit and receive polarity. (2) Input signals from the scene combine linearly within the radar and the processor prior to detection. (3) The system uses a magnitude squared method of detection. (4) Signal manipulations on the output are linear in power. The assumption of input linearity for a SAR is true outside of the radar and almost always true inside. Magnitude squared detection, one of several possible methods, is convenient for mathematical reasons and finds justification in nature as well. Following detection, additional averaging may be performed, known as incoherent integration since it occurs in the power domain, equivalently "multi-looking" in SAR parlance.

Principle 1. Coherent scene illumination

Input to a SAR is a linear summation of voltages from reflecting elements within the scene that are illuminated by a coherent EM wave field. In distinct contrast to typical optical systems, for most SARs, the illumination from each transmitted pulse is essentially monochromatic, emitted from a point source. This implies that the radiating field has structured phase fronts, which may be well represented for most purposes by spherical surfaces,

centered at the radar. Thus, the EM illumination incident on the scene is coherent. This means that the reflections from more than one scene element combine through vector addition.

Principle 2. Image power mapping

Image domain output from a SAR is a mapping of power, derived from scene reflectivity, and presented as a two-dimensional array of numbers. It is customary to think of a radar image as a visually observable entity, such as a photographic print, whose brightness (whiteness in a positive rendition) at each location is proportional to the strength of the corresponding radar echo. As was the case with the first principle, the statement seems obvious, but there are significant implications. For coherent fields, the power is given by the magnitude squared of the vector voltage sum, not by the sum of the powers of individual echoes. Constructive and destructive interference takes place between the members of the signal ensemble. This leads to substantial variations in the power estimates known as speckle, clearly seen in most SAR imagery.

Principle 3. Well-behaved transformation

The transformation from EM scene reflectivity to image data by a SAR may be characterized if and only if both the response to an isolated point scatterer and the response to a uniform Gaussian distributed scattering field are determined. This important principle applies to all partially coherent systems which can be represented with the mathematics of quadratic filters, a wide class that includes optical and radar sensors (Raney, 1969). The common feature of such systems is that they operate on an input signal field in voltage to produce an output data field in power, using linear filters and square law detection, combined according to the degree of coherence in use by the system, and by the prevailing coherence of the scene. It is well known that a linear filter may be characterized by its response to an arbitrarily short test signal, which produces its so-called impulse response function (Woodward, 1955). Application of the Fourier transform to the impulse response function produces a frequency domain description of the linear filter, known as its transfer function. It is also true that the frequency distribution of a linear filter may be determined by observation of its frequency response to white Gaussian noise. The impulse response and the system frequency transfer function are logical equivalents of each other for a linear system. This simple relationship is not sufficient for radar systems. For a SAR, as with all partially coherent systems, measurement (or specification) is needed of both the impulse response and the frequency function (Raney, 1985). The purpose of the impulse response test is to estimate the system imaging function, analogous to the point spread function encountered in optics. The purpose of the Gaussian response test is to estimate the extent of system partial coherence. The essential lesson in this principle is that a SAR is well behaved in the best sense of system theory,

even if its properties are less familiar than those of a system that is linear exclusively either in (complex) amplitude or in power (such as most optical sensors).

The following principles highlight two radiometric rules for the image radar brightness that follow from the well-behaved principle, combined with the benefits of the large time-bandwidth product that characterizes most SAR systems.

Principle 4. Conservation of energy

For any neighborhood, if the scene is a Gaussian field, and all available data are used in each case, then the tone of the corresponding area in the image is a constant, independent of system focus, system coherence, or scene coherence. A Gaussian field (mathematically speaking) is an important idealized scene for imaging studies (Goodman, 1976). A wheat field (from the point of view of a farmer) is the classic practical example, having spatial features that are far too small to be resolved by the radar and having a uniform average reflectivity over an area much larger than the resolution cell. The radar is to estimate the power reflected from the wheat field by observing the corresponding average brightness value, or tone, in the image. The Gaussian property in this case derives from the fact that for each element in the image, there are many individual scatterers, such as the heads of the wheat shafts, contributing to the total scattered signal. One might think that the tone would depend on the details of the field, including the number, size, and location of each scattering element. Even more tempting, one might think that an error in system focus, or a change in the degree of partial coherence used in the processor, would lead to a change in the image tone. Wrong on all counts! Within very broad constraints, image tone is robust even when certain parameters change, which is the point of the compact wording of this principle.

The principle of conservation of energy is very powerful. Among other things, it establishes a solid basis for radiometric calibration of a SAR whereby image brightness values can be transformed to scene reflectivity estimates. For a given SAR, conservation of energy provides assurance that image average brightness (tone) is preserved independent of the number of degrees of freedom (looks) in the image data. Furthermore, given the same viewing geometry, polarization, wavelength, and system bandwidth, two different (calibrated) radars should provide equivalent mean reflectivity estimates for the same scene. This fact allows measurements from scatterometers and other radars that are not "SAR systems" to provide reflectivity data that are directly applicable to SAR observations.

Principle 5. Conservation of confusion

For any neighborhood, if the scene is a Gaussian field, and all available data are used in each case, then the uncertainty associated with each corresponding area in the image is a constant, independent of system focus,

system coherence, or scene coherence. This principle is concerned with speckle, which is the second statistical moment of the radar brightness of a random field that has nominally constant tone. The expected uncertainty in radiometric response at each pixel usually is measured by its covariance. (The covariance often is called the variance in SAR literature, a source of potential confusion.) The covariance is defined as the square of the second moment with respect to the mean value. The normalized covariance, which is the covariance divided by the local mean squared, is a measure of the amount of averaging in the data. The statistical number of degrees of freedom (“looks” in SAR literature) is given by the mean, squared, divided by the standard deviation. Speckle radiometric variation at any image pixel is correlated with that at neighboring positions. The width of the covariance function is defined as speckle correlation length, which is equal to the inverse of the coherent bandwidth of the system (in both the range and the azimuth dimensions).

The *uncertainty* at each point is measured by the volume under the normalized two-dimensional image covariance function. The principle of conservation of confusion states that the uncertainty volume is a constant of the system, under a wide variety of circumstances. A corollary of this principle states that there is a trade-off between resolution and the amount of averaging that may be used to reduce speckle variance, which is the formal justification for the image quality of Equation 9. This consequence is an example of the uncertainty principle for energy conservative systems (Papoulis, 1968).

Principle 6. Conservation of coordinates

For a range/radar, the output mapping has constant range and azimuth coordinate locations on the imaged surface, independent of angular rotations of the sensor. This principle follows from the first and third principles, and the illumination geometry customarily used with SAR systems. It implies that there is a fundamental spatial mapping accuracy available from a range-Doppler radar such as a SAR that is inherently less sensitive to sensor pointing precision than for most other imaging systems. In range, the coordinate is proportional to time delay, which is not affected by sensor rotations. In azimuth, the coordinates are established by the frequency gradient across the scene, which in turn is determined by the vehicle along-track velocity and its instantaneous along-track position. In either dimension, an error in sensor pointing does not impact spatial location of the coordinate system with respect to the sensor.

Conclusions

SAR is a mature field, but still moving ahead. Spectacular results are now routine from such space-based systems as PALSAR, RADARSAT-2, and TerraSAR-X. There are excellent online tutorials available, including URL/SAR_Tutorial (2008). More specialized online reference materials may be found at URL/INSAR (2008) for

interferometry and URL/Polarimetry_Tutorial (2008) for polarimetry. SAR conferences and workshops worth considering include IGARSS, PolInSAR, and Fringe, among others.

Bibliography

- Attema, E. P. W., 1991. The active microwave instrument on-board the ERS-1 satellite. *Proceedings of the IEEE*, **79**(6), 791–799.
- Bamler, R., Eineder, M., Kampes, B., Runge, H., and Adam, N., 2003. SRTM and beyond: current situation and new developments in spaceborne InSAR. In *Proceedings, ISPRS Workshop on High Resolution Mapping from Space*, Hanover.
- Born, W., and Wolf, E., 1959. *Principles of Optics*. New York: Pergamon Press.
- Cloude, S. R., and Pottier, E., 1996. A review of target decomposition theorems in radar polarimetry. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(2), 498–518.
- Cloude, S. R., and Pottier, E., 1997. An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, **35**(1), 68–78.
- Curlander, J. C., and McDonough, R. N., 1991. *Synthetic Aperture Radar, Systems and Signal Processing*. New York: Wiley.
- Cutrona, L. J., Leith, E. N., Palermo, C. J., and Porcello, L. J., 1960. Optical data processing and filtering systems. *IRE Transactions on Information Theory*, **IT-6**, 386–400.
- Dubois-Fernandez, P., Angelliaume, S., Souyris, J.-C., Garestier, F., and Champion, I., 2007. The specificity of P-band POLinSAR data over vegetation. In *Proceedings POLinSAR 2007*. Frascati: European Space Agency.
- Ferretti, A., Prati, C., and Rocca, F., 2001. Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(1), 8–20.
- Freeman, A., and Raney, R. K., 2008. Improving range ambiguity performance in quad-pol SAR. In *Proceedings International Geoscience and Remote Sensing Symposium IGARSS*. Boston: IEEE.
- Gabriel, A. K., and Goldstein, R. M., 1988. Crossed orbit interferometry: theory and experimental results from SIR-B. *International Journal of Remote Sensing*, **9**(5), 857–872.
- Goodman, J. W., 1976. Some fundamental properties of speckle. *Journal of the Optical Society of America*, **66**(11), 1145–1150.
- Graham, L., 1974. Synthetic interferometers for topographic mapping. *Proceedings of the IEEE*, **62**(6), 763–768.
- Green, P. E., Jr., 1968. Radar measurements of target scattering properties. In Evans, J. V., and Hagfors, T. (eds.), *Radar Astronomy*. New York: McGraw-Hill.
- Hanssen, R., 2001. *Radar Interferometry: Data Interpretation and Error Analysis*. Dordrecht: Kluwer.
- Harger, R. O., 1974. *Synthetic Aperture Radar Systems: Theory and Design*. New York: Academic.
- Henderson, F. M., and Lewis, A. J. (eds.), 1998. *Principles and Applications of Imaging Radar*, 3rd edn. New York: Wiley. Manual of Remote Sensing.
- Jakowatz, C. V., Wahl, D. E., Eichel, P. H., Ghiglia, D. C., and Thompson, P. A., 1996. *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. Boston: Kluwer.
- Johnson, W. T. K., 1991. Magellan imaging radar mission to Venus. *Proceedings of the IEEE*, **79**(6), 777–790.
- Jordan, R. L., 1980. The Seasat-A synthetic-aperture radar system. *IEEE Journal of Oceanic Engineering*, **OE-5**, 154–164.
- Lee, J.-S., Ainsworth, T., and Chen, K.-S., 2008. Speckle filtering of dual-polarized and polarimetric SAR data based on improved sigma filter. In *Proceedings Geoscience and Remote Sensing Symposium IGARSS*. Boston: IEEE.

- Leith, E. N., 2000. The evolution of information optics. *IEEE Journal of Selected Topics in Quantum Mechanics*, **6**(6), 1297–1304.
- Leith, E. N., and Upatnieks, J., 1965. Reconstructed wavefronts and communication theory. *Journal of the Optical Society of America*, **52**, 1123–1130.
- Luscombe, A. P., 1988. Taking a Broader View: RADARSAT Adds ScanSAR to its Operation. In *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'90*. Edinburgh: IEEE.
- Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., and Rabaut, T., 1993. The displacement field of the Landers earthquake mapped by radar interferometry. *Nature*, **364**, 138–142.
- Moore, R. K., Claassen, J. P., and Lin, Y. H., 1981. Scanning spaceborne synthetic aperture radar with integrated radiometer. *IEEE Transactions on Aerospace and Electronic Systems*, **AES-17**(3), 410–421.
- North, D. O., 1963. An analysis of the factors which determines signal/noise discrimination in pulsed carrier systems. *Proceedings of the IEEE*, **51**, 1016–1027.
- O'Neill, E. L., 1963. *Introduction to Statistical Optics*. Reading, MA: Addison-Wesley.
- Palubinskas, G., and Runge, H., 2008. Change detection for traffic monitoring in TerraSAR-X imagery. In *Proceedings Geoscience and Remote Sensing Symposium IGARSS*. Boston: IEEE.
- Papoulis, A., 1968. *Systems and Transforms with Applications to Optics*. New York: McGraw-Hill.
- Pettengill, G. H., Ford, P. G., Johnson, W. T. K., Raney, R. K., and Soderblom, L. A., 1991. Magellan: radar performance and data products. *Science*, **252**, 260–265.
- Porcello, L. J., Massey, N. G., Innes, R. B., and Marks, J. M., 1976. Speckle reduction in synthetic aperture radars. *Journal of the Optical Society of America*, **66**(11), 1305–1311.
- Raney, R. K., 1969. Quadratic filter theory and partially coherent optical systems. *Journal of the Optical Society of America*, **59**(9), 1149.
- Raney, R. K., 1971. Synthetic aperture radar and moving targets. *IEEE Transaction on Aerospace and Electronic Systems*, **AES-7**(3), 499–505.
- Raney, R. K., 1985. Theory and measure of certain image norms in SAR. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-23**(3), 343–348.
- Raney, R. K., 1998. Radar fundamentals: technical perspective. In Henderson, F., and Lewis, A. (eds.), *Principles and Applications of Imaging Radar*. New York: Wiley.
- Raney, R. K., 2007. Hybrid-polarity SAR architecture. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(11), 3397–3404.
- Raney, R. K., 2008. Space-based remote sensing radars. In Skolnik, M. (ed.), *The Radar Handbook*, 3rd edn. New York: McGraw-Hill. Chapter 18.
- Raney, R. K., Luscombe, A. P., Langham, E. J., and Ahmed, S., 1991. RADARSAT. *Proceedings of the IEEE*, **79**(6), 839–849.
- Shannon, C. E., 1948. A mathematical theory of communication. *Bell System Technical Journal*, **27**(379–423), 623–656.
- Skolnik, M. (ed.), 2008. *Radar Handbook*, 3rd edn. New York: McGraw-Hill.
- Souyris, J.-C., Imbo, P., Fjortoft, R., Mingot, S., and Lee, J.-S., 2005. Compact polarimetry based on symmetry properties of geophysical media: the $\pi/4$ mode. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 634–646.
- Souyris, J.-C., and Mingot, S., 2002. Polarimetry based on one transmitting and two receiving polarizations: the $\pi/4$ mode. In *Proceedings IEEE International Geoscience and Remote Sensing Symposium*. Toronto.
- SRTM, 2005. <http://www2.jpl.nasa.gov/srtm/mission.htm>
- Stokes, G. G., 1852. On the composition and resolution of streams of polarized light from different sources. *Transactions of the Cambridge Philosophical Society*, **9**, 399–416.
- Turin, G., 1960. An introduction to matched filters. *IEEE Transactions on Information Theory*, **6**(3), 311–329.
- Urkowitz, H., 1964. The effect of antenna pattern on the performance of dual-antenna radar airborne moving target indicators. *IEEE Transactions on Aerospace and Naval Electronics*, **ANE-11**, 218–223.
- URL/ASAR, 2008. <http://envisat.esa.int/instruments/asar/>
- URL/INSAR, 2008. http://en.wikipedia.org/wiki/Interferometric_synthetic_aperture_radar
- URL/PALSAR, 2008. <http://www.eorc.jaxa.jp/ALOS/about/palsar.htm>
- URL/Polarimetry_Tutorial, 2008. <http://earth.esa.int/polsarpro/tutorial.html>
- URL/PSInSAR, 2008. <http://www.geodesy.miami.edu/sar.html>
- URL/RADARSAT, 2008. <http://www.radarsat2.info/>
- URL/SAR_Tutorial, 2008. http://www.ccrs.nrcan.gc.ca/radar/index_e.php
- van Zyl, J. J., Zebker, H. A., and Elachi, C., 1987. Imaging radar polarization signatures: theory and observation. *Radio Science*, **22**(4), 529–543.
- Venkataraman, G., Rao, Y. S., and Rao, K. S. 2005. Application of SAR interferometry for Himalayan glaciers. In *Proceedings FRINGE 2005*. Frascati: European Space Agency.
- Walker, J., 1980. Range-Doppler imaging of rotating objects. *IEEE Transactions on Aerospace and Electronic Systems*, **AES-16**(1), 23–52.
- Woodward, P. M., 1955. *Probability and Information Theory, with Applications to Radar*. New York: McGraw-Hill.
- Zebker, H., and Villasenor, J., 1992. Decorrelation in interferometric radar echoes. *IEEE Transaction on Geoscience and Remote Sensing*, **30**, 950–959.
- Zebker, H. A., and Goldstein, R. M., 1986. Topographic mapping from interferometric synthetic aperture radar observations. *Journal of Geophysical Research*, **91**(B5), 4993–4999.
- Zebker, H. A., and Lu, Y., 1998. Phase unwrapping algorithms for radar interferometry: residue-cut, least-squares, and synthesis algorithms. *Journal of the Optical Society of America*, **15**(3), 586–598.
- Zebker, H. A., and Van Zyl, J. J., 1991. Imaging radar polarimetry: a review. *Proceedings of the IEEE*, **79**(11), 1583–1606.

Cross-references

Calibration, Synthetic Aperture Radars
Data Processing, SAR Sensors
Radars
Radiation, Polarization, and Coherence
Subsidence

RADARS

Keith Raney
Applied Physics Laboratory, Johns Hopkins University,
Laurel, MD, USA

Synonyms

Active sensor; Altimeter; Scatterometer; Sounder;
Synthetic aperture radar (SAR)

Definition

RADAR, an acronym derived from *RA*dio *DE*tectioN *ANd* *R*anging, is an active sensor that transmits and receives electromagnetic (EM) waves. Remote sensing radars are designed to generate terrain maps or to measure a wide variety of geophysical parameters such as wind speed over the ocean, forest biomass, or annual sea-level rise.

Introduction

Radar was first field tested and patented in 1904 (Hulsmeyer, 1904) but did not attract serious attention at the time. Sir Watson-Watt, with his 1937 Chain Home coastal radars (20–30 kHz, 350–750 kW peak power), is generally credited (somewhat disingenuously) as the “father of radar,” due in large part to the spectacular success of this system in the Battle of Britain (Battle, 2008). Radar has evolved considerably since 1950, earning a major place in the suite of modern remote sensing instruments, which includes imaging radars (especially synthetic aperture – SAR – systems, both airborne and spaceborne), scatterometers, altimeters, and radar sounders (Raney, 2008). An outstanding early example is the Seasat satellite (Figure 1) launched in 1978, which (as its name suggests) was designed for oceanic observations. Three of its sensors were radars – a synthetic aperture radar, an altimeter, and a scatterometer. These three Seasat instruments established the initial paradigm for virtually all subsequent radars of their respective classes (Barrick and Swift, 1980; Evans et al., 2005).

Space-based synthetic aperture radar (SAR) systems capable of 1 m resolution have become the norm, with systems under development or launched by at least seven

countries. The (range) measurement precision of surface height change is on the order of 1 mm per year, as established by SARs (see “*Radar, Synthetic Aperture*”) and radar altimeters (see entry *Radar, Altimeters*). Several nations are sponsoring radars for exploration at the Moon and beyond.

Any radar system observes a scene within constraints established by the frequency, polarization, and illumination geometry of the emitted signals. After reception of an echo and subsequent processing, the observable parameters include position, reflectivity, polarization, and phase, from which geophysical information of the observed object may be retrieved. The various types of remote sensing radars are differentiated only by the design emphasis on one or more of their basic radar parameters. Military radars usually are optimized to respond to point objects (discrete targets). Such systems are characterized by fine resolution, frequency agility, moving target capability, and multiple target dynamic tracking. In contrast, the primary objective of most remote sensing radars is to measure the reflectivity of relatively large areas (distributed targets). This leads to systems characterized by moderate resolution, post-detection averaging to improve image quality, and constraints on the calibration and stability of the final image products.

Remote sensing radars (Ulaby et al., 1981) usually are classed according to their wavelength λ (related to frequency f by $\lambda f = c$, the speed of light). Most remote sensing radars use wavelengths (Table 1) that are comparable to the structural detail of many geophysical features of interest, an attribute that is a central aspect of radar data products. (International frequency allocation protocols limit the choice of frequency and usable bandwidth



Radars, Figure 1 Seasat (launched in 1978) carried a synthetic aperture radar, a scatterometer, and an altimeter, whose antennas ranged in size from 1 m (the altimeter) to 10 m (the length of the SAR antenna).

Radars, Table 1 Common radar band designations and their corresponding wavelengths and frequencies

Radar band	Wavelength (cm)	Frequency (GHz)
P	77–133	0.22–0.39
L	19–77	0.39–1.55
S	7.1–19	1.55–4.20
C	5.2–7.1	4.20–5.75
X	2.8–5.2	5.75–10.90
Ku	1.67–2.50	12–18
Ka	0.75–1.67	18–40
W	0.30–0.54	56–100

available to radars to be less than the letter designation definitions (see Table 1). Alert: “P-band” as used in radar remote sensing actually refers to the 70 cm wavelength band, thus technically at the high end of L-band. The so-called L-band remote sensing radars operate at 23 cm wavelength.) For any given system, the emitted polarization and wavelength are known constants over the scene, whereas the angle of illumination varies according to the relative position of each scatterer. For the radars of interest in this entry, the transmit antenna and the receive antenna are assumed to be in the same location, which is the monostatic configuration. (Certain radars rely on transmit and receive antennas that are separated, which is known as the bistatic configuration (Willis and Griffiths, 2007).)

The transmitting antenna determines the polarization of the emitted wave. Remote sensing radars usually use linear polarization (H or V), with the exception of Earth-based radar astronomical observatories, which invariably use circular polarization (Campbell et al., 2002), and certain meteorological radars (Bringi and Chandrasekar, 2005). As a rule, only one polarization may be transmitted at a time. Likewise, the receiving antenna selects one component of polarization of the reflected signal. If there is a change in the polarization of the transmitted wave during either reflection or propagation, then the polarization of the reflected wave will differ accordingly. This has practical implications for microwave remote sensing. The reflectivity dependence on polarization may be observed through a radar whose antenna and supporting systems are designed to handle more components and to process the data to derive polarization signatures, known in general as polarimetry (see “*Radiation, Polarization, and Coherence*” and “*Radar, Synthetic Aperture*”).

In addition to wavelength, polarization, and swath width, the radar system parameters determine the resolution and radiometric quality of the resulting image. Resolution is a measure of the radar’s sharpness, where “high resolution” means “small minimum discernable separation distance” between two closely spaced objects. “Closely spaced” is situation dependent: A good space-based SAR may offer 1 m resolution over a 5 km square image, in stark contrast to a wind-field scatterometer designed to parse the data onto a 25 km grid spanning

many hundreds of kilometers. Range resolution (along the line of sight of the radar beam) is inversely proportional to the transmitted pulse’s bandwidth. Resolution in the crossbeam direction – sometimes termed “azimuth” in certain geometries – is determined by the radar’s antenna beamwidth, unless multi-pulse signal processing schemes are applied. The radiometric quality of the radar data products usually is limited by the amount of noncoherent averaging. Such averaging, known in statistics as “degrees of freedom,” is expressed as “looks” in SAR data (see “*Radar, Synthetic Aperture*”) and as the “k factor” in scatterometry (see “*Radar, Scatterometers*”).

It often is said that radar is “all weather,” but this generalization clearly is not universally true. From space, the ionosphere and/or atmosphere may corrupt or even prevent radar wave propagation. Passage through the ionosphere may induce Faraday rotation, which is a change in the plane of polarization forced by a planet’s magnetic field. If more than a few degrees, this would degrade or even destroy the polarization properties of the transmitted and received signals (Pisacane, 2005). The Faraday rotation of a linearly polarized E-vector is proportional to $M\lambda^2$, where the rotation measure M is a function of ionospheric electron density. Clearly, Faraday rotation is more of an issue for radars at long wavelengths, such as L-band and P-band for Earth-orbital systems. The ionosphere also introduces dispersion and, under certain unfavorable circumstances, effectively cuts off propagation. Thus, for example, it was not possible for the 5 MHz MARSIS radar sounder to probe the Martian surface during daylight hours, because the cut-off frequency under those conditions increased to about 10 MHz. MARSIS was designed to work as a surface sounder during dark hours and as an ionospheric sounder during daylight hours (Gurnett, et al., 2005). The 12 cm wavelength of the Magellan Venus radar (see “*Radar, Synthetic Aperture*”) was chosen in response to the trade-off between propagation through Venus’ very dense atmosphere (for which longer wavelengths would be better) and synthetic aperture radar system considerations (for which shorter wavelengths would be better). Propagation speed is retarded along the path length from an ocean-viewing altimeter to the Earth by a very small fraction of the speed of light but sufficient nevertheless to impose range measurement errors of many meters. These errors must be estimated and compensated before the required cm-level accuracy can be achieved (see “*Radar, Altimeters*”). Atmospheric moisture is more of an issue for radars at shorter wavelengths, effectively limiting Ka-band systems, for example, under certain circumstances to a range of only a few km in terrestrial applications. In contrast, weather radars are designed specifically to observe atmospheric moisture.

Reflectivity

The information value of radar data products is conditional upon the backscatter collected by the instrument. Radar reflectivity falls naturally into two types, either

discrete (point) or diffuse (distributed), described, respectively, by radar cross section, sigma, and normalized reflectivity, sigma zero.

Sigma

Reflectivity is described by σ (sigma) if the backscatter is due to an isolated discrete scatterer – often called a point target (Skolnik, 2008). In this context, “point” usually means that the physical size of the reflecting object is smaller than the spatial resolution of the radar. A three-sided or trihedral corner reflector is a good example of a point target. Strong reflections also may occur from a simple plane surface when it is orthogonal to the line of sight of the radar. The effective area of a trihedral or dihedral is determined by the object itself and also for small flat planes. For extensive planar surfaces, such as a calm ocean surface illuminated at vertical incidence by a radar altimeter, the radar’s wavelength λ and height h above the surface determine the effective area. This is known as the (first) Fresnel zone, whose diameter is given by $(2h\lambda)^{1/2}$. The backscattered signal strength from these geometric shapes is relatively large, especially when the surface in question is smooth and is at least moderately conductive. Collectively, these are known as specular backscatter, as all the reflections tend to be in phase and directed back in the direction of the radar’s illumination. If two or more point scatterers are present in the same resolution cell of a radar, then the signals reflected simultaneously from them interfere, complicating their collective phase characteristics.

Sigma has units of area since the value of σ is equal to the surface area of the equatorial disk of a hypothetical, isotropically scattering sphere that would be large enough to cause the same reflected power observed at the radar as that seen from the target. From this equivalence, σ is known as the radar cross section (RCS). The “area” expressed by RCS (or σ) bears rather little resemblance to the size of an object; it is much more dependent on backscatter directivity due to the object’s shape and on its electrical properties. For example, at X-band, a human being and an F-18 aircraft both have (average) cross sections of about 1 m² or less, whereas a triangular corner reflector with sides 1 m long at their intersection has a cross section of about 4,500 m². Sigma may be expressed in dBm², which conveys the RCS of an object in decibels (dB) with respect to a standard RCS of 1 m². Thus, these two examples have cross sections of 0 dBm² and 36.5 dBm², respectively. (In contrast, the RCS of a stealth aircraft is on the order of 40 dBm².)

The properties of corner reflectors are reviewed extensively in Ruck et al. (1970) and summarized for remote sensing applications in Ulaby et al. (1982). As a rule of thumb, the RCS of a corner reflector or flat plate is proportional to A^2/λ^2 , where A is the effective area of the shape orthogonal to the incoming radar illumination. Carefully constructed corner reflectors are useful as calibration references.

Corner reflection also occurs in a natural situation. Dihedral reflection often is observed from the intersection between the side of an iceberg and the sea surface, or the combination of tree trunks surrounded by standing water. Trihedral reflection may occur from angular structures on a large vehicle such as a ship or from mutually orthogonal surfaces in a cluster of buildings.

Sigma zero

For distributed terrain features giving rise to diffuse scattering, the conventional description of radar reflectivity to be found in the literature is σ^0 (sigma zero, sometimes denoted sigma nought). Sigma zero represents the average reflectivity of a given surface, normalized with respect to the reflecting area on the surface, where the area is delimited by the range and cross-range resolution of the radar. Sigma zero sometimes is called the scattering coefficient (Ulaby et al., 1981), as in effect it quantifies the reflection efficiency of the terrain feature in question. It is a dimensionless fraction, usually expressed in decibels (dB) that describes the ratio of the (average) backscattered power compared to the power of the incident field. The value of σ^0 depends on the physical and electrical properties of the reflecting material and on the wavelength, polarization, and illumination’s angle of incidence. Typical values of terrain or oceanic reflectivity when viewed obliquely by a radar are much smaller than unity, on the order of –10 dB to –30 dB. In contrast, a radar such as an altimeter or (ice) sounder that looks down on a (radar-bright) horizontal surface may encounter strong reflectivities that are larger than unity, on the order of 10 dB or more. Certain large-area distributed scatterers such as the Amazon rain forest have been proven to have relatively constant normalized reflectivity of known value (see “*Calibration, Synthetic Aperture Radars*”). These are used routinely as calibration references for space-based remote sensing radars such as SARs (Luscombe et al., 2006) and scatterometers.

Sensitivity

The reflectivity observed by a radar is always in competition with noise, primarily additive receiver noise. For some types of remote sensing radars, additive noise is expressed as σ_{Neq}^0 or noise equivalent sigma zero (Ulaby et al., 1982). This number may be interpreted literally to suggest an equivalent hypothetical area whose scattering coefficient corresponds to the observed noise level, thus quantifying the inherent (additive) noise floor of the radar with respect to the span of reflection coefficients to be encountered in the radar’s application. A radar’s σ_{Neq}^0 always may be reduced (thus improved) by transmitting more power, in contrast to multiplicative noise, whose level is always proportional to the transmitted power, in the same way that the strength of the intended radar reflections is proportional to the power transmitted.

Wavelength

For a given wavelength and polarization, there are three significant factors that influence the strength of the signals reflected back toward the radar. These factors are surface roughness, local incident angle, and material dielectric constant. Surface roughness is the most important characteristic, followed by incident angle effects. Surface roughness, a qualitative norm, in the radar context always is in comparison to wavelength λ .

Typical remote sensing radars operate within a sector of the EM spectrum whose wavelengths span ~ 70 cm to less than 1 cm. Radiation at these wavelengths interacts with the detailed structural scales typical of natural and fabricated objects. The Rayleigh criterion (Rayleigh, 2008) defines a smooth surface to be one whose mean topographic height variation $h \lambda / (8 \cos \theta)$ where θ is the incident angle between the incoming ray and the normal vector of the mean surface. Radar backscatter is small from a smooth surface when seen obliquely, since most of the incoming energy is reflected in the forward specular direction, away from the radar. When illuminated at the same incidence, reflection from a rough surface will be more isotropic, reflecting a sensible portion of the EM energy back toward the radar. Within limits, greater roughness tends to increase the relative backscatter seen by the radar. Two radars that have different wavelengths may see the same object differently. Likewise, an increase in the angle of incidence may convert an extended object from “rough” to “smooth.” Of course, the Rayleigh roughness criterion is only a rule of thumb, as the point of transition between the two extremes is gradual. Nevertheless, it does provide valuable guidance on the interplay between key radar and object parameters.

The penetration of EM waves into a medium is proportional to radar wavelength and is suppressed by the material’s intrinsic dielectric losses δ . For low-loss media, such as dry sand or clean fresh ice, for which the loss tangent $\tan \delta \ll 1$, the penetration depth $D = \lambda / (\pi \tan \delta)$. This relationship is a major factor in wavelength selection for radar sounders.

Bragg scattering

There is another type of coherent scattering that occurs with certain rough surfaces having supposedly random features. If the scatterer positions are organized such that they are spaced periodically in range, then when viewed obliquely, their respective backscatters coherently reinforce each other when the several reflections are in phase. Phase agreement occurs when their range spacings r_B satisfy the Bragg scattering condition (Bragg, 2008):

$$r_B = \frac{n\lambda}{2 \sin \theta} \quad (1)$$

The relationship derives from the field of crystallography, in which the parameter n allows longer period structure to contribute to the coherent reflected component.

In microwave applications, only the lowest order case ($n = 1$) usually is dominant.

Bragg scattering, as with other forms of co-phase combination of individual reflections, leads to a total reflected power that is proportional to the square of the number of constituents, which is always greater than that from randomly phased scatterers for which the total power is proportional only to their number. Reflectivity from certain natural rough surfaces, such as the ocean, is predominantly Bragg scattering (Zebker and Van Zyl, 1991). This is a consequence of the inherent selectivity of coherent scatterer reinforcement: in an ensemble of randomly located scatterers, if there is a subset whose spacings satisfy the Bragg condition, then the probing wave will find them.

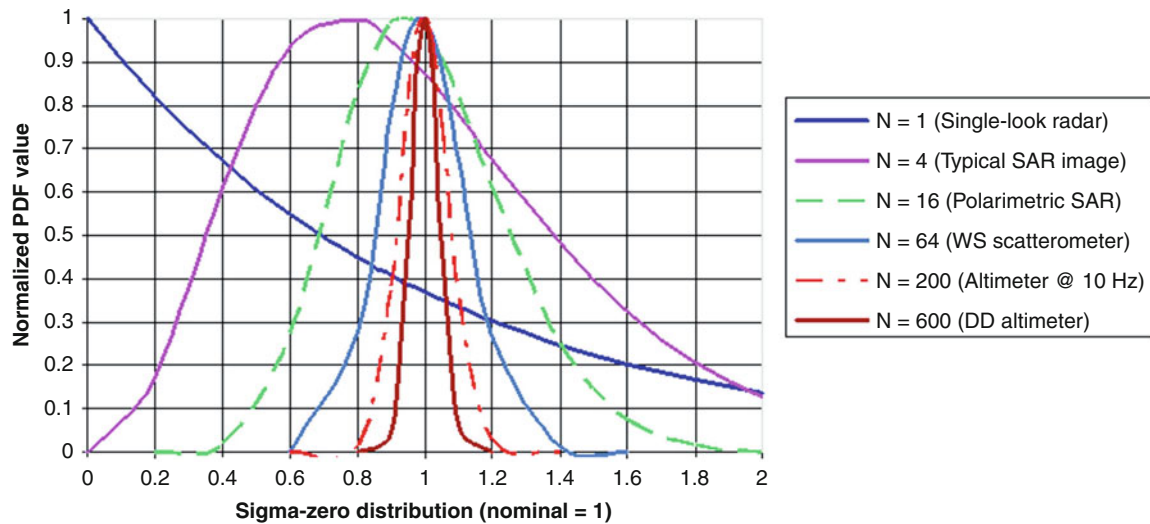
Accuracy and precision

Reflectivity measurements suggest that a direct relationship exists between σ^0 at each resolved location in the system’s data product and the radar brightness (power) derived from the scene’s echoes. Unfortunately, radars do not observe directly the power of each individual echo. As a result, the accuracy and precision of any radar measurement deserves careful attention. In the vast majority of remote sensing radars, at any given instant, many individual echoes from within a given resolution cell are presented simultaneously to the receiver. Each such backscatter constituent has its own amplitude and phase. These combine to form one signal (amplitude and phase) that represents the radar’s output at that particular resolution cell. This seemingly simple fact has many consequences; it is the smoking gun – and also the blessing in disguise – behind many subtle aspects of remote sensing radar performance. Because of this phase-sensitive summation at their front end, radars always must be respected as coherent sensors (the resulting radar measurements may or may not take advantage of signal coherence. Coherent radars do, in contrast to noncoherent radars, as reviewed in subsequent paragraphs), setting them apart from most other remote sensors.

To quote Wikipedia, “Accuracy is the degree of veracity while precision is the degree of reproducibility” (Accuracy, 2008). In other words, accuracy is the relative agreement between the average result of a measurement and the actual value of the intended parameter, whereas precision is the concentration of a distribution of measurements (described by standard deviation for example) with respect to their mean. (Alert: Standard deviation is denoted in most statistical references by the symbol σ ; potential confusion with a target’s radar cross section σ usually may be determined by context.)

Accuracy

For any remote sensing radar whose radiometric properties are stable over time, its accuracy may be “tuned up” by calibration, the objective of which is to minimize – even to drive to zero – any systematic residual difference



Radars, Figure 2 The probability distribution functions (normalized to unity maximum value) of the power estimated by a radar in response to distributed random scatterers (having $\sigma^0 = 1$) parameterized by the number of statistically independent looks. These PDFs belong to the family of gamma functions.

between the radar's mean measurement and the true value. For users, post-calibration accuracy is the radar's measurement characteristic that counts. Remote sensing radars have moved successfully into the realm of quantitative measurement, for which the norm is relatively small tolerance on acceptable error for their geophysical measurements. Effective calibration depends on preflight system characterization, calibration references built into the radar, and on external "end-to-end" references of known radar reflectivity, such as corner reflectors or large areas of distributed scatterers (see "*Calibration, Synthetic Aperture Radars*").

Precision

Most (remote sensing) radars are precision challenged when viewing distributed (diffuse) backscatter. This is the unavoidable consequence of combining within one radar resolution cell many individual echoes that have about the same frequency, but whose phase and amplitude are random. Precision reflects radar radiometric uncertainty, which in the wider world of radar is known as "fading." In SAR (see "*Radar, Synthetic Aperture*"), jargon for the same concept is "speckle."

Radar measurement precision in response to diffuse backscatter under most circumstances is well modeled by signals that are complex Gaussian. Consider the problem of measuring the radar backscatter of a scene whose normalized reflectivity is unity. This notional radar transmits one pulse, receives the backscatter, and after demodulation, amplification, and pulse compression (none of which are relevant to the present issue) forms an estimate of the received power by magnitude-squared detection. If this experiment is repeated many times, the results taken individually converge to an exponential probability

distribution function, as in Figure 2. Note that the most likely normalized reflectivity value is zero (!) for this $N = 1$ (one-look) case. The sample points are distributed widely from zero to several times the nominal value of unity, which is about as imprecise as the measurement can get. The variance (and standard deviation) of this example is 1. If the variance is viewed as noise, then this result implies for the $N = 1$ case an effective signal-to-noise ratio of 1. The alert reader will have noted that this "noise" is always proportional to the mean value of the normalized reflectivity, hence sometimes termed "multiplicative." (Single-look data have their own special value, examples of which may be found in the entry "*Radar, Synthetic Aperture*.")

Now, using the same data, average together four individual power responses to get a new estimate. Then, as shown in the figure, for the resulting $N = 4$ PDF curve, the sigma zero accuracy improves (as the mean of the PDF is closer to the actual value of 1.0), and the precision (the width of the PDF) also improves. The standard data product from most SAR systems is four-look imagery. Increasing the amount of incoherent (power-summed) averaging continues to improve measurement precision. The table cites the numbers typical of representative remote sensing radars. The examples include the following: $N = 16$, two hybrid-polarimetric lunar radars (see "*Radar, Synthetic Aperture*"); $N = 64$, typical of wind speed scatterometers (see "*Radar, Scatterometers*"); $N = 200$ and $N = 600$, representative, respectively, 1/10 s waveform averaging for a conventional noncoherent altimeter and a delay-Doppler (partially coherent) altimeter (see "*Radar, Altimeters*"). As a rule of thumb, precision (standard deviation) improves in proportion to the square root of the number of statistically independent samples in the average. Alternatively, the mean-squared-to-variance

ratio of the PDF equals the number of statistically independent constituents in the average.

Improved precision by incoherent averaging requires that two background conditions be met. First, the detected returns must be statistically independent (mutually uncorrelated). If not, then the results would be compromised by residual correlation within the ensemble of summed returns. Second, the backscatter must be wide-sense stationary from a statistical perspective. In other words, each of the individual measurements must “see” essentially the same scene during the interval over which the average is taken. These limits may be expressed more formally in expressions that are specific to each type of remote sensing radar.

Radars end to end

Traditional textbook treatments of radar systems usually start with the radar equation (Ridenour, 1947; Skolnik, 2008; Ulaby et al., 1982). The radar equation describes the peak power of the received signal in response to a single-pulse observation of an isolated, discrete scatterer. For simple radars, this equation is helpful as an introduction to the way radar performance depends on system parameters. One essential aspect of the radar equation is to quantify the performance of the system in the presence of additive noise, in terms of the observed signal-to-noise ratio SNR. However, the radar equation is just part of the story. Most remote sensing radars are comprised of two logical sections: the radar and the processor. This section closes with several ancillary comments.

The radar

The radar radiates a modulated electromagnetic (EM) frequency, often characterized by its mean wavelength, λ . For each transmission, the radar generates a pulse that is directed by the antenna’s pattern toward the scene. A fraction of the incident field is reflected back toward the radar where it is gathered by the receiving antenna. The received signal, being weak, is vulnerable to additive noise which accompanies the signal through the rest of the system.

There also are unwanted hitchhikers with the signal due to sidelobes in the antenna pattern, aliases due to sampling by the transmitted pulse sequence, analog-to-digital range sampling that may contribute quantization noise, and range ambiguities. Noise contributions of this kind are known collectively as multiplicative, since their level is always proportional to the level of the intended radar signal. This noise is quantified by the multiplicative noise ratio MNR, similar in principle to the additive noise ratio SNR. Unlike additive noise, multiplicative noise cannot be reduced by an increase in the radar’s transmitted power.

Both signal and noise are amplified in the receiver and demodulated from the carrier frequency to a low-pass or video frequency. In general, these video signals are written into memory. Radars that retain phase data for subsequent processing are known as coherent, with implications on the radar itself as well as the processor. At some stage

during processing, most radars estimate the power of the radar returns, an operation known as detection, the most common form being square law, for which the operation is magnitude squared, $|x|^2$.

After square-law detection, the radar (averaged) output is an estimate of reflectivity, with units of power. If the measurement is one dimensional, as it would be for an altimeter, then the output usually is in graphical form, with the horizontal axis proportional to time delay and the vertical axis proportional to radar reflectivity. For radars with two-dimensional output format, as is the case for air traffic control, wind-field scatterometers, or synthetic aperture radars, both spatial dimensions of the output are needed to represent radar brightness as a function of position. The single-pulse peak signal-to-noise ratio for a single target of radar cross section σ is given by:

$$SNR = \frac{P_T G^2 g^2(\theta, \varphi) \lambda^2 C_R \sigma}{(4\pi)^3 R^4 K_B T B F L} \quad (2)$$

The usefulness of a radar measurement must be judged in comparison to the relative noise level which competes with the desired signal observed through the same system. All electronic systems have inherent internal noise, due to the random motion of free electrons in conducting elements (Johnson, 1928). This receiver noise, although of extremely low level, competes with the very weak signals from radar backscatter in the front end of the radar, usually the low-noise amplifier (LNA). It is conventional in radar systems (Skolnik, 2008) to represent the average receiver noise power by $N_{rec} = K_B T B F$ where:

K_B = Boltzmann’s constant (1.38×10^{-23} J/K)

T = effective receiver temperature (Kelvin)

B = receiver noise bandwidth (Hz)

F = receiver noise factor

The receiver noise factor (or noise figure if stated in *decibels* (dB) as is customary) describes the noise performance of a given receiver compared to that of a perfect receiver. Naturally, a real receiver is always short of the ideal, so the noise factor is larger than unity. Typical receivers have noise figures of about 3 dB or somewhat more. The other three parameters describe the noise to be expected from an ideal receiver, having a given bandwidth and effective temperature. The product $K_B T B F = N_0$ is the noise per unit bandwidth (or noise spectral density) at the input to the receiver. From this stage on, all signals are amplified alike; hence, (at least to first order for the response to distributed targets) the SNR at the front end of the radar determines the SNR preserved in the output data products. In a multi-pulse radar, additive noise is statistically independent pulse to pulse.

The remaining parameters in the radar equation are determined by attributes of the radar, save for the target’s reflectivity σ . These parameters are:

P_T = (peak) transmitted power

G = gain of the antenna (one way, power)

$g(\theta, \phi)$ = one-way directivity (power) of the antenna, in azimuth and elevation, normalized
 λ = radar wavelength
 C_R = range pulse-compression ratio
 R = range from the radar to the target
 L = factor to account for losses in the radar

Peak power and compression ratio work together. Almost all remote sensing radars use some form of pulse modulation, so that more energy can be transmitted within a peak-power limit imposed by hardware constraints. For such pulses, pulse compression increases the peak power of the sharpened pulse, thus improving the SNR of point targets in proportion. Likewise, larger transmitted energy improves the average SNR of distributed reflections (characterized by σ^0) relative to the background additive receiver noise (characterized by noise equivalent sigma zero σ_{Neq}^0). The gain of the antenna usually is reciprocal, acting on both the transmitted EM field and the received field; hence, gain is squared in the round-trip radar equation. The same comment applies to the antenna's directivity. (Directivity and gain differ, in that directivity describes the shape of the antenna pattern, whereas gain includes losses or other imperfections imposed by the antenna as the microwave energy passing through it.)

The single-pulse radar equation of Equation 2 shows the classic R^{-4} dependence on range. This is due to the R^{-2} spreading loss of the radiated energy in traverse from the radar to the target and another R^{-2} spreading loss by the energy reflected back toward the radar. This range dependency depends on the radar's viewing geometry and the way in which a sequence of received pulses may be combined in processing. Side-looking radars usually are characterized by a multi-pulse radar equation with an R^{-3} range dependence (thanks to their antenna patterns whose width expands in proportion to range). Altimeters and sounders may have range dependence of R^{-2} in response to specular scattering from Fresnel-zone delimited areas on surfaces viewed orthogonally by these radars.

The smallness of the reflected signals in the radar's front end may be illustrated by the Seasat synthetic aperture radar. That SAR had an average radiated power of 55 W, about the same as the power consumed by a modest incandescent light bulb. Typical range for that radar was about 850 km, and the antenna had an area of approximately $10 \times 2 = 20 \text{ m}^2$. The effective power received by Seasat from an object having a radar cross section of 10 m^2 would be on the order of 10^{-17} W , about a factor of 1/1,000,000,000,000,000,000 smaller than the transmitted power. Even so, most Seasat imagery has very good average SNR, typified by $(-18 \text{ dB } \sigma^0 - 10 \text{ dB})$ with respect to the radar's σ_{Neq}^0 of -21 dB .

The processor

The discussion to this point has considered primarily the return from only one transmitted pulse. All radars of interest in remote sensing take advantage of processing over

many pulses. As transmissions are repeated, the resulting amplitude range lines constitute the input signal set for the processor. Cross-range (azimuth) processing depends on the relative phase relationship between pulses. In the simplest case, range processing and azimuth processing are orthogonal and may be considered as separate operations. In general, they are coupled, which must be accounted for in the processor. From a data processing point of view, the several types of radars used in remote sensing are distinguished primarily by the specific signal structures and processing strategies employed in each embodiment.

Most remote sensing radars depend on extensive signal processing. The more sophisticated systems, such as synthetic aperture radars, exploit the relative phases of a large number of individual single-pulse responses (see "*Radar, Synthetic Aperture*," and "*Data Processing, SAR Sensors*"). Simpler radars use only detection and video display in their signal processing. Signal processing includes detection and post-detection filter operations, such as data averaging, and transformation into the (Doppler) frequency domain, usually to sharpen the cross-range resolution of the system. The main message of this brief summary is that from a signal processing point of view, although remote sensing radars obey fundamentally similar rules based on common physical principles, they are distinguished by their respective processing strategies.

Pulse compression and bandwidth

Signal-to-noise performance of remote sensing radars is determined by the transmitted pulse's energy, even though the radar equation is usually stated in terms of peak power. Since transmitted energy is fundamental to a radar's performance, phase-modulated pulses much longer than the intended resolution are used so that the radar is better able to transmit a relatively large amount of energy while still satisfying a high resolution (i.e., short pulse) requirement. Following reception, the radar (or its processor) must compress the data, thus collapsing the long modulated reflection to the more familiar sharp impulse response from a point target. Range pulse modulation and compression is an excellent scheme used extensively in all classes of remote sensing radars.

The time-bandwidth product $TBP = \Delta t B_p$ is a parameter of fundamental importance for pulse-compression systems, where Δt is the length of the modulated (coded) pulse before compression and B_p is its bandwidth. Prior to pulse compression, values of $TBP > 100$ are typical in remote sensing radars and may be very much larger. Following compression, $TBP \sim 1$ since there is no remaining internal phase modulation, and the pulse width in time $\tau = 1/B_p$. The value of a pulse's TBP cannot be less than unity. This fact, analogous to the uncertainty principle of physics, is a consequence of a fundamental law of information theory (Woodward, 1955).

The theoretical resolution τ of any radar system is given by its inverse bandwidth. Interpretation of this rule in the

range direction is obvious, since the time-to-distance scaling is always by the factor $c/2$ (150 m per microsecond), where c is the speed of light, approximately 3×10^8 m/s. (Why divide by two? Because the radar's range measurement – and hence the scaling on all range-dependent parameters such as effective pulse length – is two-way which is logically equivalent to a one-way traverse at half the actual round-trip velocity.) Interpretation of the inverse bandwidth rule is situation specific, as it depends on a time-to-distance scaling that is determined by the geometry of the particular radar and viewing conditions. Note that in a perfect system, the bandwidths of the original modulated pulse and of the compressed pulse are the same. The azimuth resolution achieved through a synthetic aperture radar may be viewed as a pulse-compression operation, being described by mathematics that are essential identical to a linearly frequency-modulated pulse. In this case, the time-to-distance scaling is proportional to the radar's forward velocity rather than (half of) the speed of light.

Coherence

A coherent system is one for which signal phase structure is systematically retained so that it is available for signal processing. In comparison to an incoherent system, extra effort is required to make a radar coherent, motivated of course by the extra power phase structure affords the processor.

For the simplest form of radar, the phases of the transmitter ϕ_i^T and the local oscillator ϕ_i^{LO} are random values, uniformly distributed between zero and 2π , thus different for every i th pulse. This is characteristic of a noncoherent (sometimes “incoherent”) radar. In a noncoherent radar, each received signal following demodulation would have a reference phase ($\phi_i^{LO} - \phi_i^T$) which would be different for all returns. Given such randomness in the reference phase, there is no way to take advantage of systematic phase variations that might be inherent to the signal. In other words, it is impossible to do phase-dependent processing on data from a noncoherent radar. Conventional altimeters and laboratory scatterometers are examples of noncoherent remote sensing radars.

A given radar is coherent if, and only if, the reference phase difference ($\phi_i^{LO} - \phi_i^T$) in the post-demodulation data ensemble has the same value for all of the returns. To achieve this, the radar may be designed to lock the reference phase of both the transmitter and the receiver to a constant value. Alternatively, the local oscillator phase may be reset to match the phase of the transmitter on each and every transmission, in which case the system is said to be coherent on receive. In either case, the inter-pulse phase structure in the signal ensemble is stable, so that it may be used in processing. Synthetic aperture radar is the best known example of a coherent radar.

For many remote sensing radars, the concept of coherence is more broadly defined. If filter operations are

applied after signal detection, then the end-to-end performance of the radar becomes partially coherent (known as mixed integration in radar astronomy). This means that signal phase structure may be used in processing, but only within limits. Outside of the range of signal coherence, the phase structure is not used. Multi-look processing of SAR data is a typical example. Likewise, there may be forced departures from system coherence beyond the control of the two oscillator reference phases. These would occur, for example, if there were atmospheric phase perturbation of the propagating signal or if there were random and uncorrected motions of the radar platform. Under such conditions, the desired signal phase structure may not be fully accessible for processing. These effects lead to partial coherence of the system.

Multi-channel radars

Many remote sensing radars are designed to take advantage of systematic differences between the data streams through more than one channel. Ocean-viewing scatterometers are good examples of noncoherent multi-channel radars, as they rely on the backscattered power differences in polarization and viewing aspect to infer wind speed and direction (see “*Radars, Scatterometers*”).

Coherent multi-channel measurements take advantage of relative phase differences. For example, the interferometric mode on CryoSat uses differential phase to measure the cross-track slope of a continental ice sheet (see “*Radar, Altimeters*”). Polarimetric SARs (see “*Radar, Synthetic Aperture*”) depend on phase and amplitude differences between channels for orthogonal polarization components to measure the covariance matrix of the backscattered field (coherent dual polarization) or the complex scattering matrix (full or quadrature polarization).

In a more generalized sense of “multi-channel,” mutually coherent radar data may be gathered over a given scene on two time-separated occasions. If there are changes as small as a fraction of a wavelength between radar observations, these motions can be observed by differential phase processing (see “*Radar, Synthetic Aperture*,” and “*Microwave Radiometers, Interferometers*”). There are two cases: (1) time changes and (2) elevation variations. (1) If the two viewing positions along their respective radar passes are identical, then any observed differences are due to motion within the scene, either to moving objects such as vehicles or to very slow geophysical processes, such as land subsidence. (The optimum time interval between such measurements depends of course on the rate of change in the scene and the wavelength of the radar.) (2) If the two viewing positions of the radar are offset in the plane orthogonal to the line of sight to the scene, then the resulting phase differences may be interpreted as elevations in the scene.

Examples such as these illustrate the unique power of certain radar remote sensing systems to make measurements on the scale of their wavelength, regardless of their spatial resolution.

On space-based radars

Unlike an airborne platform that can go anywhere at any time (subject to fuel and air-space limitations), a satellite's position and velocity when in orbit about a planetary body are rigidly governed by orbital dynamics, summarized compactly by Kepler's laws. Further, access by a space-based remote sensing radar to a given area of interest depends on the rotation rate of the planet as well as the satellite's position along its orbit and the radar's viewing geometry.

The performance of radars is conditional upon the velocity of their host platform far more sensitively than other types of sensors. This is especially true for radars that use any form of range-Doppler processing, such as altimeters, sounders, or SARs. Hence, the performance of a range-Doppler radar depends to first order on the velocity and range of its host spacecraft, which in turn depends on where in the universe that spacecraft may be. The velocity V of a spacecraft in orbit at altitude h above a planet of radius R and mass M is given by:

$$V = [MG/(R + h)]^{1/2} \quad (3)$$

where G is the universal gravity constant (N is the standard symbol for Newton, force, with the units m kg s^{-2}) $6.67 \times 10^{-11} \text{ Nm}^2 \text{ kg}^{-2}$. Table 2 lists representative spacecraft velocities for bodies in the solar system that have been visited, or are likely to be observed, by range-Doppler radars. Feasible satellite altitudes are limited below by the prevailing atmospheric density. The final column of the table lists the altitude-velocity product hV corresponding to each entry. This product is a scaling factor that characterizes the range-Doppler space confronting a SBR intended to be deployed in that environment. There is approximately a 40-fold spread in the value of this parameter, from the Earth to Jupiter's moon Europa. It follows that radar designs that work in one situation may not be at all appropriate if migrated to a different planetary body.

Radars, Table 2 Spacecraft velocities and typical range-Doppler scaling parameters for a variety of planets and moons relevant to remote sensing radars (such as SARs, sounders, and altimeters), either already visited (e.g., the Moon, Mars, and Venus) or are interesting candidates

Body	Mass (kg)	Radius (km)	Altitude h (km)	V (m/s)	hV (km ² /s)
Earth	5.97 E + 24	6,380	800	7,466	6,000
Venus	4.87 E + 24	6,052	300	7,151	2,200
Mars	6.4 E + 23	3,397	400	3,353	1,600
Titan	1.35 E + 23	2,575	200	1,801	360
Ganymede	1.4 E + 23	2,631	100	1,849	185
Calisto	1.08 E + 23	2,400	100	1,697	170
Moon	7.35 E + 22	1,737	100	1,634	160
Europa	4.8 E + 22	1,569	100	1,385	140
Enceladus	1.08 E + 20	504	100	109	11

Thumbnails of remote sensing radars

Synthetic aperture radar

In its most general form, an imaging radar is a device designed to provide a two-dimensional portrayal of the radar backscatter returning from the field illuminated in range and azimuth. Space-based microwave imagers are synthetic aperture radars (with the exception of certain early Soviet ocean-observing real-aperture systems). As with all imaging systems, SAR image products are rated according to their resolution, where "higher" is "better." Higher resolution always implies wider bandwidth, in both range and azimuth. Azimuth bandwidth derives from the Doppler signatures set up by the motion of the radar with respect to the illuminated field. Resolution by itself is not sufficient to determine the image quality of importance to most applications. SAR images are degraded by a multiplicative "self-noise" known as speckle, which is a direct consequence of the coherence required by the radar-processor combination to form the synthetic aperture and the resulting enhanced resolution. Speckle can be reduced only through supplemental incoherent processing, known in SAR jargon as multi-looking. Additional looks require proportionally more bandwidth. It follows that two-dimensional bandwidth (range and azimuth) is the driving requirement for this class of radar.

Space-based SAR systems have motivated fruitful specializations in quantitative applications in a wide variety of areas, comprehensively reviewed in *The Manual of Remote Sensing* (Henderson and Lewis, 1998). For an introduction to topics such as SpotSAR, ScanSAR, polarimetry, and interferometry, see "[Radar, Synthetic Aperture](#)."

Radar altimeters

The main objective of a space-based altimeter is to measure the distance between the radar and the surface and to deduce geophysical quantities of the surface from the reflected waveforms. Relatively small changes (on the order of cm) in mean sea surface height may correspond to substantial differences in the corresponding geophysical parameters. Sea surface height measurements have become essential for a wide variety of applications in oceanography, geodesy, geophysics, and climatology (Fu and Cazanave, 2001). With the exception of near-polar ice, Earth-orbiting oceanographic altimeters have seen relatively little application over nonaquatic surfaces.

A satellite-based altimeter systematically circles the Earth, gathering waveforms along track. The resulting data provide measurements of sea surface height, significant wave height, and wind speed. Although one might consider altimeters to be relatively simple one-dimensional (range measurement) instruments, their phenomenal accuracy and precision requires elegant microwave implementation and innovative signal processing. Since the mid-twenty-first century, coherent and interferometric radar techniques are being incorporated into innovative altimeter systems (see "[Radar, Altimeters](#)").

Scatterometers

Space-based remote sensing scatterometers measure the normalized backscatter with sufficient precision and accuracy to deduce the value of one or more parameters of geophysical significance. For example, the power reflected from the ocean back to a radar is a function of surface roughness at the scale of the radar's wavelength, which in turn is a function of the local wind (Quilfen et al., 2004). Estimation of wind speed and direction over the open ocean is the most common application for these instruments. In addition to open ocean observations, calibrated data from this class of remote sensing radar have been applied to a variety of large-area surface features, such as determination of sea ice coverage, mapping the boundaries between the principal ice zones of Greenland, or global estimation of tropical deforestation. In all such applications, the emphasis is on measurement of mean reflectivity over large areas, rather than mapping fine spatial detail. These radars typically have resolution no better than several kilometers, supported over swaths of 1,000 km or more.

In the ocean application, relatively small changes in radar backscatter may correspond to substantial differences in the retrieved wind information (Kerkmann, 1998; Moore and Fung, 1979). It follows that the dominant requirement for this class of radar is the accuracy and precision of the received power measurement. Since the geophysical interpretation of scatterometric data often depends on distinguishing between two similar values of σ^0 , the results depend critically on reducing the uncertainty in the estimated value as well as getting the average value right. However, the next step, vector wind retrievals – transforming the radar backscattered power into accurate estimates of wind speed and direction – is far from trivial. Indeed, the technique fails for the very low wind speeds that do not generate wavelength-scale surface roughness. In the limit, any side-looking radar will generate virtually no backscatter from the sea surface in the absence of wind-driven waves, even if there is a substantial swell in the region. For more on these and related topics, see “[Radar, Scatterometers](#).”

Radar sounders

The term “sounder,” usually associated with acoustic echo sounding, derives from many centuries of oceanic shipborne depth measurements using lead lines and the like. The extension from acoustic to electromagnetic methodology is a small step in comparison. In its most general form, a radar sounder is an active device designed to transmit pulses that penetrate the volume of a target medium, from which the waveform of the resulting backscatter indicates variations in dielectric contrasts as a function of distance. Remote sensing radar sounders may be used to probe the ionosphere, the atmosphere, or solid media such as ice sheets or planetary surfaces. (A system deployed on the surface such as ground-penetrating radar (GPR) is a closely related case.)

As a sounder passes over an illuminated region, the sequence of ranging waveforms generates a profile, which is a two-dimensional reflectivity cross section of the surveyed volume. Penetration depth in general increases with wavelength and also with radiated power. On the other hand, reflectivity depends on the dielectric contrasts between internal layers; a material's dielectric constant is also a function of wavelength. It follows that space-based radar sounders must choose a frequency and bandwidth that balance the often conflicting requirements of penetration, reflectivity, and resolution, under the constraints of available power and antenna aperture. For an overview of active (radar) sounders, see Raney (2008). Alert: Radar sounders should not be confused with certain spaceborne atmospheric sounders that are passive devices.

Conclusions

Radar is a vast subject, now spanning more than 100 years; drawing upon physical principles some of which go back more than 200 years. For comprehensive coverage of imaging radar applications and an introduction to theory, see Henderson and Lewis (1998). For a somewhat more technical radar discussion, the book to consult is Toomay and Hennen (2004). For an introduction to hardcore radar design issues, see Skolnik (2008). For highlights and an historical perspective into space-based remote sensing radars, see Raney (2008).

Bibliography

- Accuracy, 2008. <http://en.wikipedia.org/wiki/Accuracy>.
- Barrick, D. E., and Swift, C. T., 1980. The Seasat microwave instruments in historical perspective. *IEEE Journal of Oceanic Engineering*, **OE-5**(2), 74–79.
- Battle, 2008. <http://www.bbc.co.uk/dna/h2g2/A733565>
- Bragg, 2008. http://en.wikipedia.org/wiki/Bragg's_Law
- Bringi, V. N., and Chandrasekar, V., 2005. *Polarimetric Doppler Weather Radar*. New York: Cambridge University Press.
- Campbell, D. B., Hudson, R. S., and Margot, J.-L., 2002. Advances in planetary radar astronomy. In Stone, R. (ed.), *Review of Radio Science*. Oxford: U.R.S.I, pp. 1999–2002. Chap. 35.
- Evans, D. L., Alpers, W., Cazenave, A., Elachi, C., Farr, T., Glackin, D., Holt, B., Jones, L., Liu, W. T., McCandless, W., Menard, Y., Moore, R., and Njoku, E., 2005. Seasat – a 25-year legacy of success. *Remote Sensing of Environment*, **94**, 384–404.
- Fu, L.-L., and Cazanave, A. (eds.), 2001. *Satellite Altimetry and the Earth Sciences*. San Diego: Academic.
- Gurnett, D. A., Kirchner, D. L., Huff, R. L., Morgan, D. D., Persoon, A. M., Duru, F., Nielsen, E., Safaeinili, A., Plaut, J. J., and Picardi, G., 2005. Radar soundings of the ionosphere of Mars. *Science*, **310**, 1929–1933.
- Henderson, F. M., and Lewis, A. J. (eds.), 1998. *Manual of Remote Sensing: Principles and Applications of Imaging Radar*, 3rd edn. New York: Wiley.
- Hulsmeyer, C., 1904. The Telemobiloscope, *Electrical Magazine*, London, Vol 2, p. 388.
- Johnson, J. B., 1928. Thermal agitation of electricity in conductors. *Physical Review*, **32**, 97–109.
- Kerkmann, J., 1998. *Review on Scatterometer Winds*. Darmstadt: European Organization for the Exploitation of Meteorological Satellites EUMETSAT.

- Luscombe, A., Thompson, A., James, P., and Fox, P., 2006. Calibration techniques for the RADARSAT-2 SAR system. In *Proceedings of EUSAR 2006*. Dresden: VDE Verlag.
- Moore, R. K., and Fung, A. K., 1979. Radar determination of winds at sea. *Proceedings of the IEEE*, **67**, 1504–1521.
- Pisacane, V. J., 2005. *Fundamentals of Space Systems*. New York: Oxford University Press.
- Quilfen, Y., Chapron, B., Collard, F., and Vandemark, D., 2004. Relationship between ERS scatterometer measurement and integrated wind and wave parameters. *Journal of Atmospheric and Oceanic Technology*, **21**(2), 368–373.
- Raney, R. K., 2008. Space-based remote sensing radars. In Skolnik, M. (ed.), *The Radar Handbook*, 3rd edn. New York: McGraw-Hill. Chap. 18.
- Rayleigh, 2008. [http://en.wikipedia.org/wiki/Gloss_\(material_appearance\)](http://en.wikipedia.org/wiki/Gloss_(material_appearance))
- Ridenour, L., 1947. *Radar System Engineering*. New York: McGraw-Hill.
- Ruck, G. T., Barrick, D. E., Stuart, W. D., and Krichbaum, C. K., 1970. *Radar Cross Section Handbook*. New York: Plenum, Vol. 1 and 2.
- Skolnik, M. (ed.), 2008. *Radar Handbook*, 3rd edn. New York: McGraw-Hill.
- Toomay, J. C., and Hannen, P. J., 2004. *Radar Principles for the Non-Specialist*, 3rd edn. Raleigh, NC: SciTech.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing: Active and Passive*. Reading, MA: Addison-Wesley, Vol. I.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1982. *Microwave Remote Sensing: Active and Passive*. Reading, MA: Addison-Wesley, Vol. II.
- Willis, N. J., and Griffiths, H. D. (eds.), 2007. *Advances in Bistatic Radar*. Raleigh: SciTech.
- Woodward, P. M., 1955. *Probability and Information Theory, with Applications to Radar*. New York: McGraw-Hill.
- Zebker, H. A., and Van Zyl, J. J., 1991. Imaging radar polarimetry: a review. *Proceedings of the IEEE*, **79**(11), 1583–1606.

Cross-references

Calibration, Scatterometers
 Calibration, Synthetic Aperture Radars
 Data Processing, SAR Sensors
 Electromagnetic Theory and Wave Propagation
 Microwave Surface Scattering and Emission
 Radar, Altimeters
 Radar, Scatterometers
 Radar, Synthetic Aperture

RADIATION (NATURAL) WITHIN THE EARTH'S ENVIRONMENT

Anthony England
 College of Engineering, University of Michigan,
 Ann Arbor, MI, USA

Synonyms

Brightness; Earth radiance; Planck spectrum; Thermal emission

Definition

Natural electromagnetic radiation within the Earth's environment is a product of natural physical processes and can be used as a remotely sensed indicator of the conditions governing those processes and/or the media through which the radiant signal passes.

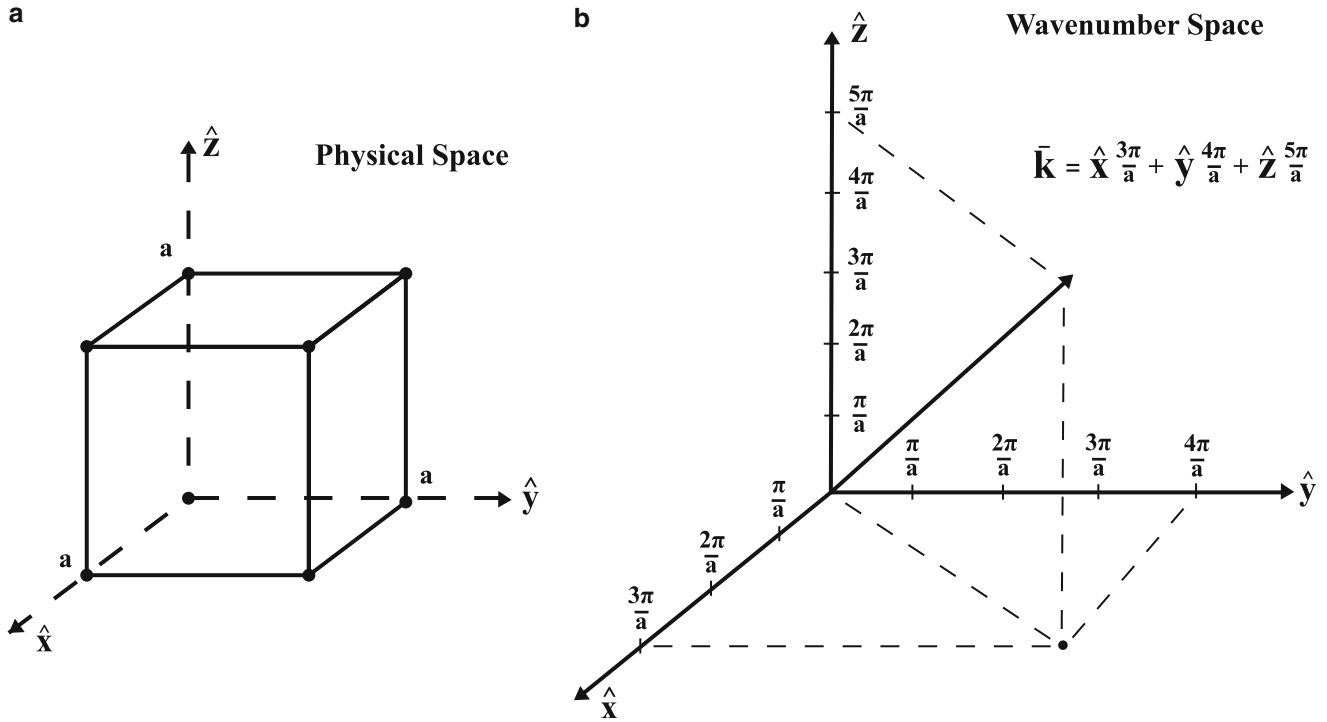
Natural radiation within the Earth's environment

The natural sources of electromagnetic radiation within the Earth's environment that are used for remote sensing can be conveniently classified by frequency. At frequencies below ~ 1 GHz, dominant sources are nonthermal and include lightning (sferics) and ionospheric synchrotron radiation. At frequencies between ~ 1 GHz (~ 30 cm wavelength, the lower end of the microwave spectrum) and 40 THz (~ 8 μm wavelength, the higher end of the thermal infrared (TIR) spectrum), natural sources are overwhelmingly the result of thermal processes. Beginning with near-infrared frequencies (or reflectance infrared frequencies), through visible frequencies, and to the upper limit of the UV frequencies, the Sun serves as the primary source for natural source, remote sensing techniques. Above UV frequencies, nuclear decay is a source of the gamma radiation used in well logging and in airborne sensing of Snow Water Equivalent (SWE). This section concerns thermal sources within the Earth's environment, i.e., natural sources at frequencies between 1 GHz and 40 THz.

Thermal radiation from elements of the Earth system that might be sensed from aircraft or spacecraft is a product of thermal energy of media at the Earth's surface or within the atmosphere. Although the span of relevant temperatures is nearly 100 K centered on ~ 275 K, it is convenient to think of the Earth as a 300 K source. The radiant brightness of a 300 K source decreases rapidly to insignificance at wavelengths shorter than 8 μm and is masked by nonthermal sources at frequencies below ~ 1 GHz. Atmospheric opacity at wavelengths between ~ 14 μm and ~ 1 mm effectively divides the thermal radiance spectrum into two classes of remote sensing techniques identified by their wavelength, TIR (8–14 μm) techniques and thermal microwave (1 mm to 30 cm) techniques. Concepts discussed in this section will be applicable to both ends of this 300 K thermal radiant spectrum and will be couched in the hybrid language of optical physics (e.g., Bohren and Huffman, 1983) and radio astronomy (e.g., Kraus, 1966).

The section is organized around six topics cast as assertions:

- (i) Planck radiation is a consequence of local thermal equilibrium.
- (ii) Brightness temperature is invariant along a ray traversing a nonabsorbing, scatterer-free medium.
- (iii) Emissive power equaling absorption is a condition of local thermal equilibrium.
- (iv) EM constitutive properties are most intuitive when expressed in the frequency domain.



Radiation (Natural) Within the Earth's Environment, Figure 1 Cartesian representations of a cubic resonant cavity whose edges are length a in physical space (a). For the condition that surfaces of the cube are Perfect Electrical Conductors (PEC), allowed standing waves are prescribed by all positive integers, p, q, r , and the wave vector, $\vec{k} = \hat{x} \frac{\pi p}{a} + \hat{y} \frac{\pi q}{a} + \hat{z} \frac{\pi r}{a}$ in wave number space (b). The wave vectors, $\vec{k}$, are normal to the corresponding wave fronts (planes of constant phase).

- (v) Real and imaginary components of frequency domain EM constitutive properties are related through Kramers–Kronig relationships.
- (vi) Emissive power is a consequence of the imaginary component of relative permittivity.
- (i) Planck radiation is a consequence of local thermal equilibrium

Thermal radiation is the thermal energy in a medium that is partitioned into radiant energy. Planck radiation is thermal radiation from a region that is in local thermal equilibrium. Planck's explanation of the process has not been improved upon (Planck, 1909). An abbreviated version is presented here because familiarity with the assumptions allows us to recognize where the assertion might break down.

Consider a dielectric cube whose edges have length a (Figure 1a), where a is much larger than any relevant electromagnetic wavelength, λ . For mathematical convenience, assume that the electromagnetic constitutive properties of the medium are locally isotropic. This limitation is readily removed once the expression for Planck, or "blackbody," radiance and its dependence upon the index of refraction is established.

Begin by identifying the electromagnetic modes that can be supported within this cube. Because $a \gg \lambda$, it is reasonable to assume that energy differences associated with the choice of boundary conditions for the cube's

electromagnetic modes are small with respect to the total energy of electromagnetic modes in volume v . For convenience, assign the boundary condition of a Perfect Electrical Conductor (PEC), so there can be no electric fields on the faces of the dielectric cube. Standing waves in the orthogonal $\hat{x}$, $\hat{y}$, and $\hat{z}$ directions shown in Figure 1a have spatial frequencies, or wave numbers, k_x , k_y , and k_z , that satisfy

$$\begin{aligned} k_x a &= \pi p \\ k_y a &= \pi q \\ k_z a &= \pi r, \end{aligned} \quad (1)$$

where p, q , and r are any positive integers. In wave number space, illustrated in Figure 1b, every allowed wave vector, $\vec{k}$, in the dielectric cube can be expressed:

$$\begin{aligned} \vec{k} &= \hat{x}(\pm k_x) + \hat{y}(\pm k_y) + \hat{z}(\pm k_z) \\ &= \hat{k} k, \end{aligned} \quad (2)$$

where $\hat{k}$ is a unit vector normal to the wave front and wave number, k , is

$$k = \sqrt{\left(\frac{\pi p}{a}\right)^2 + \left(\frac{\pi q}{a}\right)^2 + \left(\frac{\pi r}{a}\right)^2}. \quad (3)$$

Standing wave modes outside the first octant of Figure 1b are degenerate in that all distinguishable modes can be described by choosing the “+” options in Equation 2. The elemental volume, Δ_k , in wave number space associated with each standing wave mode is

$$\Delta_k = \left(\frac{\pi}{a}\right)^3 = \frac{\pi^3}{v}. \quad (4)$$

If $N(k) dk$ is the number of standing wave modes in the first octant of a spherical shell of thickness dk centered on the origin in Figure 1b, then

$$\begin{aligned} N(k) dk &= \left(\frac{1}{8}\right) \left(\frac{4\pi k^2 dk}{\Delta_k}\right) (2) \\ &= \left(\frac{k^2 v dk}{\pi^2}\right), \end{aligned} \quad (5)$$

where the “1/8” in the first parenthesis on the right of the first line designates the first octant of the spherical shell, the numerator in the second parenthesis is the volume of the spherical shell, the denominator is the volume associated with one mode, and the “2” in the third parenthesis represents the two possible polarizations of each standing wave mode.

Planck’s contribution, in addition to this formulation, was the assertion that the energy in each mode was quantized and could be expressed $m h f$ where m is an integer, h is Planck’s constant (6.6262×10^{-34} J s), and f is frequency. From Boltzmann and statistical mechanics (e.g., McQuarrie, 2000), the probability density of energy $m h f$ for an electromagnetic mode that is in thermal equilibrium with all other thermally excited modes, including non-electromagnetic modes, is

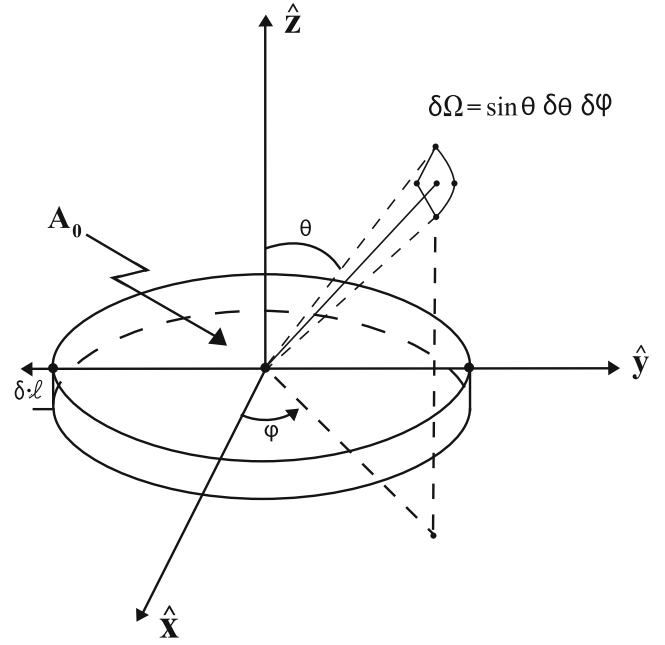
$$p(mhf) = C e^{\frac{-mhf}{k_B T}}, \quad (6)$$

where C is a normalizing constant, k_B is Boltzmann’s constant (1.3806×10^{-23} J K⁻¹), and T is temperature in Kelvin. The expected energy, $\langle U(f) \rangle$, in a particular electromagnetic mode having frequency, f , is then

$$\begin{aligned} \langle U(f) \rangle &= \frac{\sum_{m=0}^{\infty} m h f p(m h f)}{\sum_{m=0}^{\infty} p(m h f)} \\ &= \frac{hf}{e^{\frac{hf}{k_B T}} - 1} \text{ J.} \end{aligned} \quad (7)$$

If $N(f)$ is the number of modes in the cavity having frequency f , then cavity’s energy density at frequency f at thermal equilibrium becomes

$$w(f) = \frac{N(f) \langle U(f) \rangle}{v} \text{ J}^{-1} \text{ m}^{-3} \text{ Hz}^{-1}. \quad (8)$$



Radiation (Natural) Within the Earth’s Environment,

Figure 2 A pill-shaped volume within the resonant cavity of Figure 1. The faces of the pill (normal to the $\hat{z}$ axis) have area A_o , and the thickness of the pill is $\delta\ell$. Radiation escapes the pill in solid angle, $\delta\Omega$, integrated over the 4π solid angle of the sphere. The assumption is that the pill is sufficiently thin that the radiation escaping through its edge is negligible.

The independent variable in Equation 5 is k , while that in Equation 4 is f . The number of modes in terms of f equivalent to those in terms of k is obtained from the dispersion relation, $k = 2\pi f \left(\frac{n}{c}\right)$, where n is the refractive index. That is,

$$\begin{aligned} N(f) df &= N(k) dk \\ &= \frac{k^2 v}{\pi^2} dk \\ &= \left(\frac{2\pi f n}{c}\right)^2 \frac{v}{\pi^2} \left(\frac{2\pi n}{c}\right) df \\ &= \left(\frac{8\pi f^2 n^3 v}{c^3}\right) df. \end{aligned} \quad (9)$$

Equations 7 and 9 allow the energy density at thermal equilibrium, Equation 8, to be written

$$w(f) = 8\pi h \left(\frac{fn}{c}\right)^3 \left(\frac{1}{e^{\frac{hf}{k_B T}} - 1}\right) \text{ J}^{-1} \text{ m}^{-3} \text{ Hz}^{-1}. \quad (10)$$

This can be converted to spectral brightness using the concept of radiation escaping from a pill-shaped volume having thickness, $\delta\ell$, and area, A_o (Figure 2). The spectral energy in the pill $\delta W(f)$ is

$$\begin{aligned}\delta W(f) &= w(f) A_o \delta \ell \\ &= 8\pi h \left(\frac{f n}{c} \right)^3 \left(\frac{1}{e^{k_B T} - 1} \right) A_o \delta \ell \quad \text{J}^{-1} \text{ Hz}^{-1}.\end{aligned}\quad (11)$$

A strategy for identifying the Planck brightness spectrum is to equate the electromagnetic energy at frequency f stored as standing waves in the pill-shaped volume to the same energy as it leaves the volume as radiant energy. For sufficiently small $\delta \ell$, effectively all radiant energy leaves the pill-shaped volume through the faces, each having area A_o . That is,

$$\delta W(f) = 2 \int_{2\pi} \left\{ \left(A_o \cos \theta B_{fp}^b(\hat{k}) \right) \left(\frac{\delta \ell}{\cos \theta c} \right) 2 \right\} d\Omega \quad \text{J}^{-1} \text{ Hz}^{-1}.\quad (12)$$

The first “2” to the right of the equal sign in Equation 12 represents the two faces of the pill-shaped volume. The quantity inside the first parenthesis of the curly brackets represents the irradiance passing through area A_o , from solid angle $d\Omega$ centered on unit vector $\hat{k}$, where θ is the angle between $\hat{k}$ and the normal to surface A_o (Figure 2). The quantity in the second parenthesis is the time it takes for all radiant energy traveling in direction $\hat{k}$ to leave the pill-shaped volume. The final “2” in Equation 12 refers to the two possible polarizations. The “blackbody” spectral brightness in direction $\hat{k}$ is $B_{fp}^b(\hat{k})$, where superscript b indicates that this is a Planck, or “blackbody,” radiance, and the subscripts designate spectral brightness at frequency f and polarization p . The units of $B_{fp}^b(\hat{k})$ are $\text{W m}^{-2} \text{ sr}^{-1} \text{ Hz}^{-1}$.

The dielectric in the pill-shaped volume is isotropic so that neither the refractive index, n , nor the spectral brightness, $B_{fp}^b(\hat{k})$, can depend upon $\hat{k}$. This allows the curly brackets of Equation 12 to be moved outside the integral. The integral becomes 2π , and

$$\delta W(f) = 8\pi B_{fp}^b A_o \delta \ell \left(\frac{n}{c} \right) \quad \text{J}^{-1} \text{ Hz}^{-1}.\quad (13)$$

Equating the right-hand sides of Equations 12 and 13 and solving for brightness yield

$$B_{fp}^b = h f \left(\frac{f n}{c} \right)^2 \left(\frac{1}{e^{k_B T} - 1} \right) \quad \text{W}^{-1} \text{ m}^{-2} \text{ sr}^{-1} \text{ Hz}^{-1}.\quad (14)$$

Equation 14 is Planck’s equation for single polarization thermal brightness in an isotropic medium, which was our objective. We next examine the assertion and assumptions used to obtain Planck’s equation.

The assertion of local thermal equilibrium. Local thermal equilibrium means that all mechanical and

electromagnetic modes of a dielectric in a local region of a size on the order of the electromagnetic wavelength are in thermal equilibrium. Because energy will leak from a local region to regions having different temperatures more rapidly by some modes than by others, local thermal equilibrium requires local mechanisms for relatively rapid exchanges of energy between local modes if they are to remain in thermal equilibrium. Primary mechanisms for this exchange include nonlinearity in the mechanical constitutive relations and dielectric loss expressed in the dielectric constitutive relation. Nonlinearity redistributes energy among electromagnetic modes, and loss converts energy between electromagnetic and mechanical modes. This author is not aware of any situation where a natural dielectric in the Earth’s thermal environment fails to exhibit local thermal equilibrium. There might be situations in deep space where temperatures approach those of the cosmic background, 2.7 K, where the assumption of local thermal equilibrium fails.

The assumption that $\lambda \ll a$. This is a requirement that local regions of a dielectric be homogeneous and large with respect to electromagnetic wavelengths of interest. The assumption is often violated at microwave frequencies in the exchange of radiant energy among small heterogeneities, for example, water droplets in a cloud. In such situations, Rayleigh or Mie scattering theory must be combined with the concepts of radiative transfer and thermal equilibrium (e.g., Chandrasekhar, 1960; Van de Hulst, 1981; Bohren and Huffman, 1983).

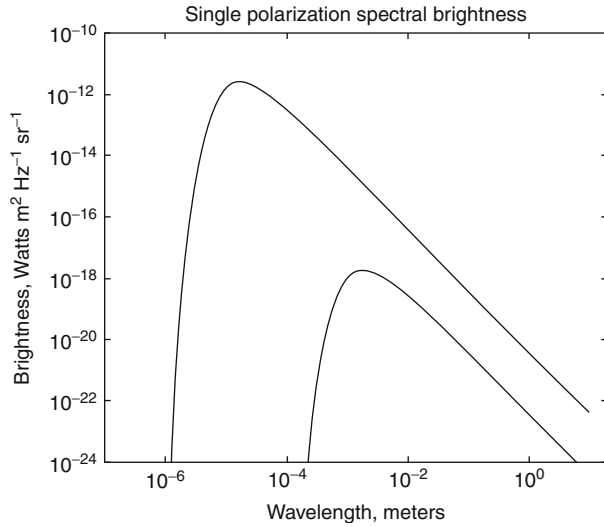
The assumption that refractive index n is isotropic. This assumption has no effect upon the spectral brightness sensed by a radiometer outside a beamfilling “black” emitter. We will see later in this section that spectral brightness normalized by the square of the refractive index is invariant upon transition from one medium to another if the interface is conditioned to avoid reflections. Consequently, Equation 14 requires that spectral brightness within an anisotropic dielectric also be anisotropic.

The assumption that $hf \ll k_B T$. Known as the Rayleigh–Jeans approximation, this assumption was neither needed nor used in our development of the Planck spectral brightness. It is included here because its use is common in the characterization of thermal microwave brightness within the Earth’s environment. With this approximation, Planck spectral brightness, Equation 14, becomes

$$\begin{aligned}B_{fp}^b &= \left(\frac{f n}{c} \right)^2 k_B T \\ &= \frac{k_B T}{\lambda^2} \quad \text{W}^{-1} \text{ m}^{-2} \text{ sr}^{-1} \text{ Hz}^{-1}.\end{aligned}\quad (15)$$

The Rayleigh–Jeans approximation offers the simple relationships that brightness varies linearly with temperature and inversely with wavelength squared.

Limitations of the Rayleigh–Jeans approximation are illustrated in log–log plots of brightness versus

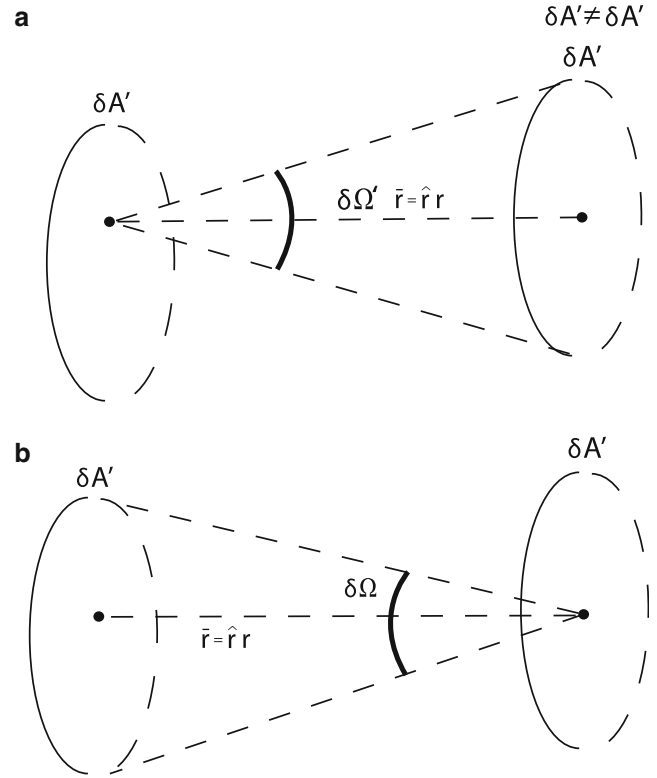


Radiation (Natural) Within the Earth's Environment, Figure 3 Planck radiation for 300 and 2.7 K "black" emitters. In log-log plots of radiance such as these, the -2 slope of longer wavelength radiation corresponds to the region of validity of the Rayleigh-Jeans approximation. Note that the approximation becomes increasingly and non-negligibly poor at wavelengths shorter than 0.1 mm in the Earth's environment and at wavelengths shorter than 1 cm for cosmic emission.

wavelength for temperatures of 300 and 2.7 K, characteristic temperatures of the Earth's environment and of dust in deep space, respectively (Figure 3). The Rayleigh-Jeans approximation is valid in wavelength regions where brightness in Figure 3 exhibits a -2 spectral slope. For a 300 K environment, the approximation begins to fail for wavelengths shorter than ~ 0.1 mm. At the 2.7 K temperature of the cosmic background, the approximation begins to fail for wavelengths shorter than ~ 1 cm. While all natural thermal sources in the Earth's environment exhibit a -2 slope over the microwave spectrum, the low temperature of the cosmic background means that cosmic background-based calibrations of satellite microwave radiometers for wavelengths shorter than ~ 1 cm should be based upon Equation 14 rather than Equation 15 if radiometer calibration is to take maximum advantage of available radiometer linearity.

- (ii) Brightness temperature is invariant along a ray traversing a nonabsorbing, scatterer-free medium.

Thermal radiant sources are always extended sources and should not be modeled as point sources. An extended source model enables use of convenient aspects of the theory of rays and brightness. Think of a ray as radiant energy traveling in a tubular path, for example, between the two surfaces, δA and $\delta A'$, shown in Figure 4. Rays occupy an incremental volume, and radiant brightness along rays in media having constant refractive indices is invariant in the absence of absorption and scattering. This invariance is readily demonstrated by comparing the flow



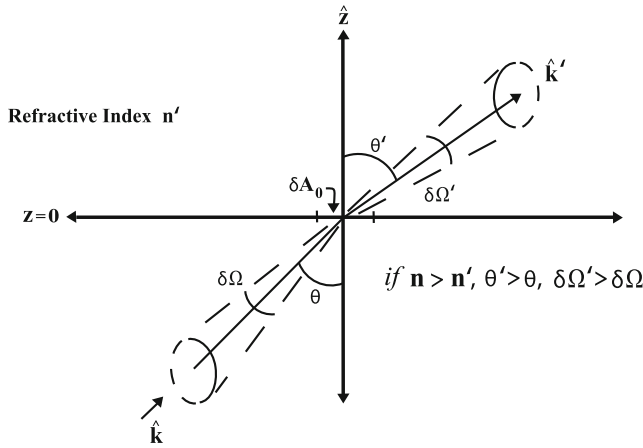
Radiation (Natural) Within the Earth's Environment, Figure 4 A demonstration of the invariance of brightness along a ray in the absence of emission, absorption, and scattering. The radiance of a ray passing through surface δA as observed at surface $\delta A'$ in (a) is identical to the radiance of the ray passing through surface $\delta A'$ in (b).

of radiant energy between the two incremental areas from the perspective of δA (Figure 4a) and from the perspective of $\delta A'$ (Figure 4b). The areas are parallel and their respective centers, designated by points P and P' , are joined by a vector, $\vec{r} = \hat{r}r$, normal to the planes of the incremental areas, where $\hat{r}$ is the unit vector in the direction of $\vec{r}$ and r is the length of the vector.

The spectral radiation pattern, or spectral brightness, of a unit area centered on point P in area δA is $B_{fp}(\hat{k})$ $\text{W m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}$, where f is frequency, p is choice of polarization, and $\hat{k}$ is the unit vector indicating radiation direction. If the transmitting medium is lossless, free of scatterers, and has a constant index of refraction, the spectral power leaving area δA and impinging upon area $\delta A'$ is (Figure 4a)

$$\delta W_{fp}(\hat{r}) = B_{fp}(\hat{r}) \delta A \delta \Omega' \quad \text{W Hz}^{-1}, \quad (16)$$

where $\delta \Omega'$ is the solid angle of $\delta A'$ as seen from P and $\hat{k} = \hat{r}$. Similarly, if the radiant brightness crossing surface $\delta A'$ is $B'_{fp}(\hat{r})$, the spectral power arriving on surface $\delta A'$ from δA is (Figure 4b)



Radiation (Natural) Within the Earth's Environment,

Figure 5 A ray passing through a dielectric interface at $z = 0$. If there are no reflections at the interface, the brightness of the ray in media of index n is related to the brightness of the ray in media n' by Equation 24.

$$\delta W'_{fp}(\hat{r}) = B'_{fp}(\hat{r}) \delta A' \delta \Omega \quad \text{W Hz}^{-1}, \quad (17)$$

where $\delta \Omega$ is the solid angle of δA as seen from $\mathbf{P}'$. Conservation of energy requires that $\delta W_{fp}(\hat{r}) = \delta W'_{fp}(\hat{r})$. Because $\delta \Omega = \frac{\delta A}{r^2}$ and $\delta \Omega' = \frac{\delta A'}{r'^2}$,

$$B_{fp}(\hat{r}) = B'_{fp}(\hat{r}). \quad (18)$$

This invariance allows radiant brightness to be viewed as a characteristic of an emitting medium without consideration of the distance between observer and medium. For example, the near-vacuum of space is effectively a lossless and scatterer-free medium with a constant index of refraction. From the perspective of a spacecraft anywhere within the solar system, the Sun is equally bright and only its apparent size varies with the inverse square of distance.

Consider lossless and scatterer-free medium, but, unlike the previous case, the refractive index of this medium varies with position. If the variation is slow over distances that are long relative to radiant wavelengths of interest, or if abrupt transitions are accompanied by interface impedance matching that avoids reflections, how does radiant brightness vary across such a transition? The question can be answered by considering the specific case of a ray traveling across a planar, nonreflecting interface. The more general case of a gradual transition can be modeled as a large set of small, discrete transitions.

An upwelling ray traversing from the lower medium to the upper medium in Figure 5 impinges at an incidence angle θ on a lossless, isotropic interface whose thickness and refractive index allow it to serve as a matching transformer between the media so that there are no reflections of the ray at the interface. $B_{fp}(\hat{k})$ is the upwelling

spectral brightness of the ray in the lower medium, and $B'_{fp}(\hat{k}')$ is the corresponding upwelling spectral brightness after it passes through the interface. The incident plane of the ray is defined as the normal to the interface and the unit vector, $\hat{k}$. For isotropic media, symmetry requires that $\hat{k}'$ also lies in the incident plane. Incident angles, θ and θ' , are related by Snell's Law,

$$n \sin \theta = n' \sin \theta', \quad (19)$$

a consequence of the requirement that phases match across the interface at every point on the interface. If δA_o is an incremental area defined as the intersection of the interface and the ray passing through it, then the spectral power, $\delta W_{fp}(\hat{k})$, in the solid angle, $\delta \Omega$, along direction $\hat{k}$ in the lower medium, impinging upon δA_o is

$$\delta W_{fp}(\hat{k}) = B_{fp}(\hat{k}) \delta A_o \cos \theta \delta \Omega \quad \text{W Hz}^{-1}. \quad (20)$$

Similarly the spectral power leaving δA_o in the solid angle $\delta \Omega'$ along $\hat{k}'$ in the upper medium is

$$\delta W'_{fp}(\hat{k}') = B'_{fp}(\hat{k}') \delta A_o \cos \theta' \delta \Omega' \quad \text{W Hz}^{-1}. \quad (21)$$

Conservation of energy requires that $\delta W_{fp}(\hat{k}) = \delta W'_{fp}(\hat{k}')$ and, canceling δA_o , yields

$$B_{fp}(\hat{k}) \cos \theta \delta \Omega = B'_{fp}(\hat{k}') \cos \theta' \delta \Omega' \quad \text{W m}^{-2} \text{Hz}^{-1}. \quad (22)$$

Expressed in polar coordinates, the incremental solid angles are $\delta \Omega = \sin \theta \delta \theta \delta \phi$ and $\delta \Omega' = \sin \theta' \delta \theta' \delta \phi'$, where $\delta \phi$ and $\delta \phi'$ are azimuth ray widths below and above the interface, respectively. Similar to the symmetry argument that $\hat{k}$ and $\hat{k}'$ lie in a common incident plane, $\delta \phi = \delta \phi'$ so that Equation 22 becomes

$$B_{fp}(\hat{k}) \cos \theta \sin \theta \delta \theta = B'_{fp}(\hat{k}') \cos \theta' \sin \theta' \delta \theta'. \quad (23)$$

Substituting Snell's Law (Equation 19) and the differential form of Snell's Law, $n \cos \theta \delta \theta = n' \cos \theta' \delta \theta'$, into Equation 23 yields

$$\frac{B_{fp}(\hat{k})}{n^2} = \frac{B'_{fp}(\hat{k}')}{n'^2} \quad \text{W m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}. \quad (24)$$

That is, spectral brightness normalized by the refractive index squared is invariant along a ray in the absence of absorption and scattering. This equivalence suggests a convention long advocated by this author of using normalized spectral brightness, $B_{fp}^n(\hat{k})$, defined

$$B_{fp}^n(\hat{k}) = \frac{B_{fp}(\hat{k})}{n^2} \quad \text{W m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}, \quad (25)$$

as the relevant parameter in a radiative transfer equation (RTE), a heuristic description of the change in brightness along a ray resulting from absorption and self-emission (e.g., Chandrasekhar, 1960). A convenient form for the RTE of a ray in local thermal equilibrium in a medium whose index varied sufficiently slowly would be

$$\frac{dB_{fp}^n(\hat{k})}{ds} + \kappa_{fa} B_{fp}^n(\hat{k}) = \alpha_f J_{fp}^n(\hat{k}) \quad (26)$$

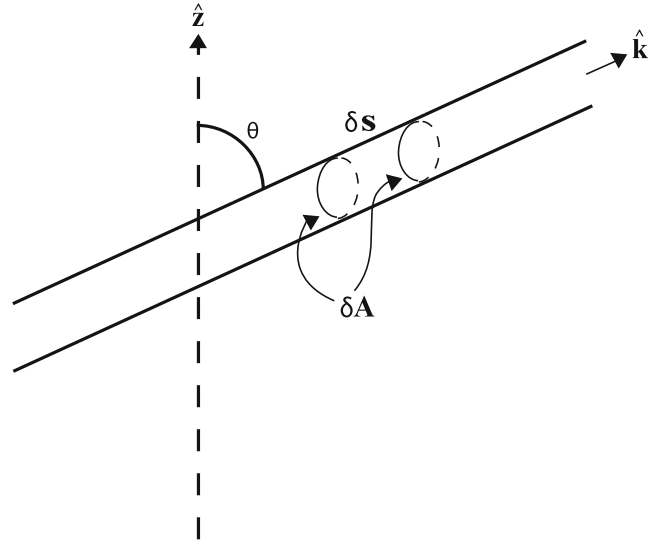
where κ_{fa} is the spectral absorption coefficient in m^{-1} at frequency f , α_f is the volume emissivity in m^{-1} at frequency f , ds is incremental distance in direction $\hat{s}$ having dimension m^{-1} , and $J_{fp}^n(\hat{k})$ is a normalized source function in direction $\hat{k}$ having units $\text{W m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}$. Normalization of the brightness of a ray traversing an ice–air interface would be significant, but normalization of the brightness of a ray traversing variations in the density of the Earth's atmosphere would be slight. For example, a ray traversing the atmosphere from sea level, where the index is 1.0003, to an Earth satellite would experience a divergence of only 0.06 %, an amount less than the likely calibration accuracy of a satellite radiometer. Normalization of the brightness of a ray traversing some part of the atmosphere of a gaseous planet, like Jupiter, could be significant.

The Rayleigh–Jeans approximation enables a more intuitive microwave parameter than normalized spectral brightness for thermal emitters in the Earth's environment. The linear relationship between brightness and temperature invites the definition of microwave brightness temperature:

$$Tb_{fp}(\hat{k}) = \frac{B_{fp}(\hat{k})}{\left(\frac{f n}{c}\right)^2 k_B} \quad \text{K sr}^{-1}. \quad (27)$$

While the supplementary SI unit, steradian, could be dropped, and most authors choose to do so, this author prefers to retain the unit as a reminder that brightness temperature has associated with it an incremental solid angle in direction $\hat{k}$. Brightness temperature is intuitive in that it is closely related to physical temperature. For example, the microwave brightness temperature of a “black” emitter having a thermal temperature of 270 K is 270 K sr^{-1} , and the brightness temperature of the night sky at 30 GHz might be 60 K sr^{-1} rather than the lessintuitive $8.3 \times 10^{-20} \text{ W m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}$. Because brightness temperature is effectively normalized by the square of the refractive index, $Tb_{fp}(\hat{k})$ is automatically invariant in the absence of absorption and scattering.

(iii) Emissive power equals absorption as a condition of local thermal equilibrium.



Radiation (Natural) Within the Earth's Environment,

Figure 6 Schematic of a ray in a one-dimensional medium. An electromagnetic ray in a medium whose properties vary only with height, z , can be described as a tube of cross section δA containing an electromagnetic wave traveling in direction $\hat{k}$ specified by its direction cosine, $\mu = \cos \theta$, with respect to $\hat{z}$. An incremental volume of the tube would be $\delta A \delta s$.

Radiative transfer in the Earth's environment is often conveniently cast as a one-dimensional problem because the physical properties of many relevant natural media are horizontally stratified. Furthermore, because electromagnetic properties are often locally isotropic, the combination of horizontal stratification and local isotropy permits a reduction of the space and direction variables from five to two, i.e., z , the distance from the surface of the Earth along zenith direction, and μ , the direction cosine with respect to the zenith direction, $\hat{z}$. For purposes of this section, media are non-scattering, i.e., all relevant constitutive properties vary slowly with z relative to the wavelength of radiant energy. The RTE for a ray of brightness temperature traveling along the path, δs , ($= \frac{\delta z}{\mu}$) in Figure 6 is

$$\mu \frac{dTb_{fp}(z, \mu)}{dz} = \kappa_{fa}(z) Tb_{fp}(z, \mu) = \kappa_{fa}(z) J_{fp}(z) \quad (28)$$

where $\kappa_{fa}(z)$ is the absorption coefficient in units of m^{-1} at frequency f and $J_{fp}(z)$ is a source function at height z , frequency f , and polarization p . The emissive power of the incremental volume, $\delta s \delta A$, i.e., the right-hand side of Equation 28, is intrinsic to the medium. The source function, $J_{fp}(z)$, is independent of direction, μ , for an isotropic medium. A further assumption of this formulation is that the refractive index of the medium does not vary sufficiently with height, z , to require a z dependence

of the direction cosine, μ , i.e., the ray does not bend appreciably.

If the medium is infinite in all directions and has a uniform thermal temperature, T , local thermal equilibrium requires that

$$\frac{d T b_{fp}(z, \mu)}{dz} = 0. \quad (29)$$

In this case of an extended region in thermal equilibrium, the brightness temperature would not have location or directional dependence even though the medium need not be uniform in z . This argument is a consequence of assertion (ii) and an extension of the principle of detailed balance. Effectively, there cannot be a flow of energy if the potential difference is zero. The Planck brightness temperature would be the same for all z and all μ . Because the source function is an intrinsic function of the medium, the combination of Equations 28 and 29 illustrates the assertion that emissive power must equal absorption.

Removing the constraint of uniform temperature so that temperature varies with height, z , the source brightness temperature must have amplitude T and be isotropic, i.e.,

$$J_{fp}(z, \mu) = T(z)u(\mu) \text{ K sr}^{-1} \quad (30)$$

where thermal temperature at height z is $T(z)$ in Kelvin, and ray direction is indicated by $u(\mu)$, which has an amplitude of unity and units of sr^{-1} . The revised RTE becomes

$$\mu \frac{d T b_{fp}(z, \mu)}{dz} + \kappa_{fa}(z) T b_{fp}(z, \mu) = \kappa_{fa}(z) T(z) u(\mu). \quad (31)$$

The terms on the right of the equal sign are collectively referred to as the emissive power of the medium in direction $u(\mu)$. Equation 31 is a first-order, linear differential equation with emissive power as a source. The integral form is

$$T b_{fp}(z, \mu) = \left(\int_{-\infty}^z \left[\frac{1}{\mu} e^{\frac{-\kappa_{fa}(z')(z-z')}{\mu}} \right] \kappa_{fa}(z') T(z') dz' \right) u(\mu) \text{ K sr}^{-1}, \quad (32)$$

where the quantity in the square brackets is the relevant Green's function. In both the differential and integral forms of the RTE, the absorption coefficient, $\kappa_{fa}(z')$, plays a key role in governing the emissive power of sources in the Earth's environment. The remainder of the section is devoted to generating plausible physical models for electromagnetic absorption at microwave and thermal infrared frequencies and linking this absorption coefficient to the electromagnetic constitutive properties of natural media.

(iv) EM constitutive properties are most intuitive when expressed in the frequency domain.

Assertion (iv) can be understood by considering the alternative, a constitutive relationship in the time domain. Consider a charge-neutral, nonionized gas, but the approach could apply equally well to charge-neutral liquids or solids. The electric displacement, $\bar{D}(t)$, is the charge distortion of a dielectric medium caused by an imposed electric field, $\bar{E}(t)$. The relationship can be expressed as

$$\begin{aligned} \bar{D}(t) &= \varepsilon(t) \otimes \bar{E}(t) \\ &= \varepsilon_o \varepsilon_r(t) \otimes \bar{E}(t), \end{aligned} \quad (33)$$

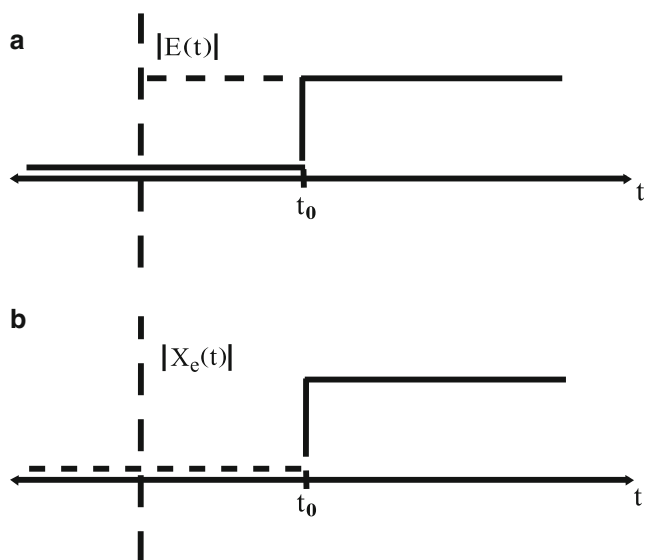
where $\varepsilon(t)$ is electric permittivity in the time domain, ε_o is vacuum electric permittivity, $\varepsilon_r(t)$ is time-dependent increase in electric permittivity relative to a vacuum, or relative permittivity, and the explicit location dependence of each of the parameters is suppressed for clarity. The relative permittivity is time dependent because the electric displacement is a consequence of reorientation or distortion of gas molecules in response to the applied electric field, and inertia requires that this reorientation or distortion not be instantaneous. Specifically, electric displacement in the time domain is a convolution of all past electric fields experienced by the gas. The symbol $\otimes$ designates the convolution.

The several contributions of gas molecules to electric displacement can be isolated by introducing the definitions of a time-dependent electric polarization, $\bar{P}(t)$, where

$$\begin{aligned} \bar{P}(t) &= \bar{D}(t) - \varepsilon_o \bar{E}(t) \\ &= \varepsilon_o (\varepsilon_r(t) - 1) \otimes \bar{E}(t) \\ &= \varepsilon_o X_e(t) \otimes \bar{E}(t). \end{aligned} \quad (34)$$

The quantity $X_e(t) = \varepsilon_r(t) - 1$ is the electric susceptibility in the time domain. The effect upon a gas of neutral atoms caused by a step in the applied electric field at time t_o is depicted schematically in Figure 7. The distortion of the electron field about the nucleus of an atom occurs with characteristic times on the order of 10^{-15} s, a period that is short relative to the periods of thermal radiance in the Earth's environment. From this perspective, the adjustment of the electron field about an atom can be considered as effectively instantaneous as depicted in Figure 7b.

A gas comprised of molecules of more than one atom behaves in a more interesting manner. If an imposed electric field causes an induced polarization that is not aligned with the electric field, or if a molecule has a permanent electric moment, for example, the water molecule, an imposed electric field can cause the molecule to rotate, absorbing and reradiating energy. Similarly, an imposed electric field can cause molecular bonds between atoms to stretch or bend absorbing and reradiating energy. These processes involve acceleration of one or more atoms. Unlike the adjustment of an electron field about an atom, accelerations of atoms in more complex



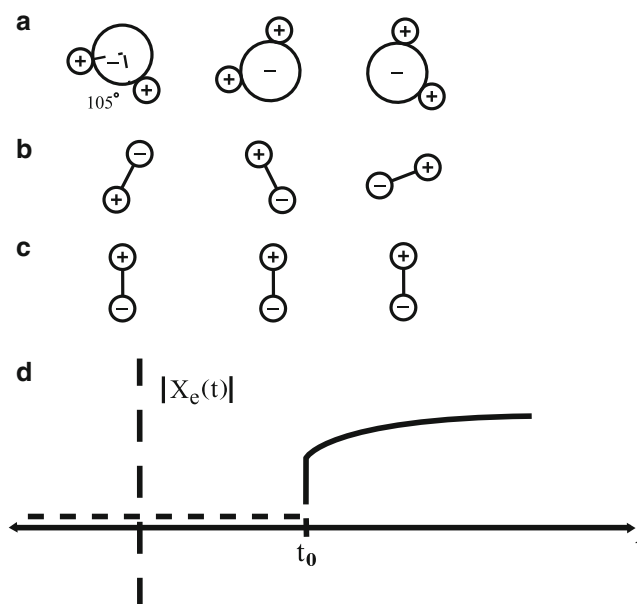
Radiation (Natural) Within the Earth's Environment,

Figure 7 The dielectric response of an isolated atom to an imposed electric field. The electric displacement of electrons about an atomic nucleus of an isolated atom occurs rapidly ($\sim 10^{-15}$ s). From the perspective of the periods of thermal radiation in the Earth's environment, this adjustment is instantaneous.

molecules are not instantaneous relative to the periods of thermal radiance. Through processes of rotating or distorting complex molecules, absorption or emission of thermal radiant energy often results in spectral signatures that are diagnostic of the composition of the gas, its temperature, and its pressure. The Debye and Lorentz models illustrate these processes. More comprehensive discussions of spectral signatures can be found in texts on molecular spectroscopy.

The Debye process

The water molecule is arguably the most interesting environmental constituent from the perspective of remote sensing because it is central to weather and climate and because the water molecule exhibits an inherent electric dipole moment that is unique among abundant natural materials (Debye, 1929). The water molecule can be pictured as a large oxygen atom bonded to two small hydrogen atoms. The angle between lines joining the center of the oxygen atom to the centers of the hydrogen atoms is 105° as depicted in Figure 8a. The bond between the hydrogen and oxygen atoms is primarily covalent, which ordinarily results in an absence of electric moments typically found in ionic bonds. The hydrogen–oxygen bond in the water molecule is an exception. The bonding electrons are heuristically described as spending more time around the oxygen atom than around the hydrogen



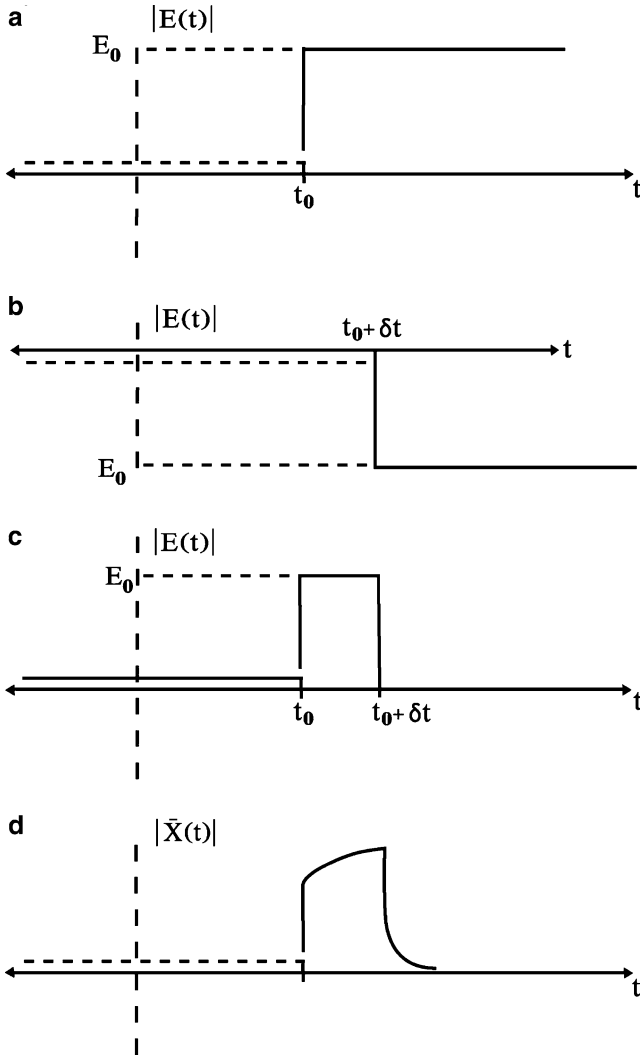
Radiation (Natural) Within the Earth's Environment,

Figure 8 Schematic of the electric displacement response of water vapor molecules to a step electromagnetic field. Water vapor molecules have permanent electric moments (a) that can be modeled as an array of electric dipoles (b). Imposition of an electric field at time t_0 will cause free dipoles to spin about electric field lines (c), but the adjustment to a step electric field will take time relative to the periods of thermal radiation in the Earth's environment. Following the instantaneous adjustment of the electrons about each nucleus, represented by an immediate rise in electric susceptibility, the adjustment of the electric dipoles to the steady-state field is represented in (d) as an asymptotic approach to a steady-state dielectric susceptibility.

atom causing a time-average permanent electric moment that is modeled as the electric dipole shown in Figure 8b. An oscillating electric field of the right frequency will couple with the permanent electric moment causing the molecule to rotate, or spin, at discrete frequencies dictated by quantum mechanics. The process can be understood without appealing to quantum mechanics through use of the Debye model (1929).

If the experiment of a uniform step electric field, $\bar{E}_o$, is imposed upon a water vapor, the result will be an instantaneous adjustment of the electron fields around each atom as described for a monatomic gas and a much slower rotation of water molecules so that their electric moments align counter to the imposed electric field as shown in Figure 8c. The amplitude of the consequent electric susceptibility is depicted in Figure 8d where the average of the time needed for the rotational adjustment of molecules is described by the characteristic time, τ . For any time $t > t_o$, the electric polarization is modeled:

$$\bar{P}(t - t_o) = \epsilon_o(\epsilon_i - 1)\bar{E}_o + \epsilon_o(\epsilon_f - \epsilon_i)\left(1 - e^{-\frac{(t-t_o)}{\tau}}\right)\bar{E}_o, \quad (35)$$



Radiation (Natural) Within the Earth's Environment,

Figure 9 The electric displacement response of water vapor molecules to an impulse electromagnetic field. An electric field impulse in a linear system is represented as a positive step at time t_0 (a), followed by a negative step of equal amplitude at time $t_0 + \delta t$ (b). The consequent electric susceptibility of water vapor is then the superposition of positive and negative response like that in Figure 8. The susceptibility at any time t (c) is the sum of the step functions in Figures 9a and 9b. This impulse becomes for forcing function applied to the dielectric (d) is dependent upon the electric fields at all earlier times so that electric polarization, $\bar{P}(t)$, becomes a convolution of electric susceptibility and the imposed electric field.

where $\varepsilon_o(\varepsilon_i - 1)\bar{E}_o$ is the initial electric polarization resulting from the adjustment of the electron fields around atoms and $\varepsilon_o(\varepsilon_i - 1)\bar{E}_o$ is additional electric polarization from the rotation of the water molecules. The electric polarization as $t \rightarrow \infty$ is $\varepsilon_o(\varepsilon_f - 1)\bar{E}_o$.

An impulse, the forcing element of a convolution, is created by adding a negative step, $-\bar{E}_o$, at time $t + \delta t$ as

shown in Figure 9. The consequent amplitude of the electric susceptibility is shown in Figure 9d. For time $t > t_0 + \delta t$, the electric polarization resulting from an impulse at time t_0 is

$$\begin{aligned} \delta \bar{P}(t - t_0) &= \varepsilon_o(\varepsilon_f - \varepsilon_i) \left(e^{-\frac{-(t-t_0+\delta t)}{\tau}} - e^{-\frac{-(t-t_0)}{\tau}} \right) \bar{E}_o \\ &= \varepsilon_o(\varepsilon_f - \varepsilon_i) \left(\frac{\partial}{\partial t_0} \left\{ e^{-\frac{-(t-t_0)}{\tau}} \right\} \delta t_0 \right) \bar{E}_o \\ &= \varepsilon_o(\varepsilon_f - \varepsilon_i) \left(\frac{e^{-\frac{-(t-t_0)}{\tau}}}{\tau} \right) \bar{E}_o \delta t_0. \end{aligned} \quad (36)$$

Given this impulse response of the water vapor system, the electric polarization at time, t , resulting from a history of imposed electric fields, $\bar{E}(t_0)$, where $t_0 < t$, is

$$\bar{P}(t) = \varepsilon_o(\varepsilon_i - 1) \bar{E}(t) + \int_{-\infty}^t \varepsilon_o(\varepsilon_f - \varepsilon_i) \frac{e^{-\frac{-(t-t_0)}{\tau}}}{\tau} \bar{E}(t_0) dt_0. \quad (37)$$

Equation 37 in the frequency domain becomes

$$\begin{aligned} \bar{P}(\omega) &= F[\bar{P}(t)] \\ &= \varepsilon_o(\varepsilon_i - 1) \bar{E}(\omega) + \varepsilon_o(\varepsilon_f - \varepsilon_i) \\ &\quad F \left[\int_{-\infty}^t \varepsilon_o(\varepsilon_f - \varepsilon_i) \frac{e^{-\frac{-(t-t_0)}{\tau}}}{\tau} \bar{E}(t_0) dt_0 \right] \end{aligned} \quad (38)$$

where $F[f(t)]$ is the complex Fourier transform of $f(t)$ and $\bar{E}(\omega)$ is the imposed electric field in the frequency domain. The complex Fourier transform in Equation 38 is

$$\begin{aligned} F[\dots] &= \frac{1}{\sqrt{2\pi}} \int_{t=-\infty}^{t=\infty} \left\{ \int_{t_0=-\infty}^{t_0=t} \varepsilon_o(\varepsilon_f - \varepsilon_i) \frac{e^{-\frac{-(t-t_0)}{\tau}}}{\tau} \bar{E}(t_0) dt_0 \right\} e^{-j\omega t} dt \\ &= \frac{\varepsilon_o(\varepsilon_f - \varepsilon_i)}{\sqrt{2\pi}} \int_{t_0=-\infty}^{\infty} e^{\frac{t_0}{\tau}} \left\{ \int_{t=t_0}^{\infty} \frac{e^{-\frac{-(t-t_0)}{\tau}}}{\tau} dt \right\} \bar{E}(t_0) dt_0 \\ &= \frac{\varepsilon_o(\varepsilon_f - \varepsilon_i)}{\sqrt{2\pi}} \int_{t_0=-\infty}^{\infty} e^{\frac{t_0}{\tau}} \left\{ \frac{e^{-\frac{t_0}{\tau}}}{1+j\omega\tau} \right\} \bar{E}(t_0) dt_0 \\ &= \frac{\varepsilon_o(\varepsilon_f - \varepsilon_i)}{1+j\omega\tau} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-j\omega t_0} \bar{E}(t_0) dt_0 \right\} \\ &= \frac{\varepsilon_o(\varepsilon_f - \varepsilon_i)}{1+j\omega\tau} \bar{E}(\omega). \end{aligned} \quad (39)$$

Equations 38 and 39 yield the electric polarization in the frequency domain:

$$\bar{P}(\omega) = \varepsilon_o \left\{ (\varepsilon_i - 1) + \frac{(\varepsilon_f - \varepsilon_i)}{(1 + j\omega\tau)} \right\} \bar{E}(\omega). \quad (40)$$

Unlike electric polarization in the time domain, which involves a convolution, electric polarization in the frequency domain is a simple product of a complex function of frequency and the electric field in the frequency domain. While a transformation from the time

domain to the frequency domain always converts a convolution to a simple product, demonstrating this characteristic for electric permittivity emphasizes the intuitive power in this case of working in the frequency domain.

The electric polarization in the frequency domain is defined

$$\begin{aligned}\bar{P}(\omega) &= \bar{D}(\omega) - \epsilon_0 \bar{E}(\omega) \\ &= \epsilon_0 (\tilde{\epsilon}_r(\omega) - 1) \bar{E}(\omega),\end{aligned}\quad (41)$$

where $\bar{D}(\omega)$ is the electric displacement in the frequency domain and the complex function, $\tilde{\epsilon}_r(\omega)$, is the relative electric permittivity in the frequency domain. Equations 40 and 41 yield the relative permittivity:

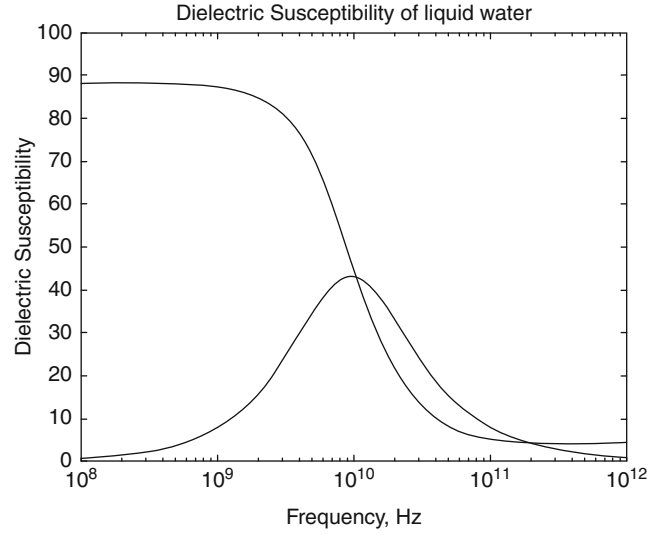
$$\tilde{\epsilon}_r(\omega) = \epsilon_i + \frac{\epsilon_f - \epsilon_i}{1 + j\omega\tau}, \quad (42)$$

which is the frequency domain expression for relative permittivity of a Debye relaxation process, i.e., a process of reorientation.

The mathematical description of the Debye relaxation process is most easily understood by introducing the relaxation frequency, $f_o = (2\pi\tau)^{-1}$, and comparing the frequency dependences of the real and imaginary parts of the relative permittivity. For $\tilde{\epsilon}_r = \epsilon'_r - j\epsilon''_r$,

$$\begin{aligned}\epsilon'_r &= \frac{\epsilon_f + \epsilon_i \left(\frac{f}{f_o}\right)^2}{1 + \left(\frac{f}{f_o}\right)^2} \\ \epsilon''_r &= \frac{(\epsilon_f - \epsilon_i) \left(\frac{f}{f_o}\right)}{1 + \left(\frac{f}{f_o}\right)^2}.\end{aligned}\quad (43)$$

The dielectric loss will be shown to be a direct result of the imaginary component of the complex relative permittivity, and the real and imaginary parts of relative permittivity will be shown to be interdependent – that is, one can be obtained from extensive knowledge of the other. The imaginary component of relative permittivity in the Debye relaxation process (Equation 43) is zero both at zero frequency and at infinite frequency, and maximum loss will occur at frequency $f = f_o$. The f_o of water vapor is 22.235 GHz (Liebe, 1969; Waters, 1976; Ulaby et al., 1981). The f_o of liquid water varies between 9 GHz at 0 °C and 17 GHz at 20 °C (Grant et al., 1957; Stogryn, 1971). The ϵ_i and ϵ_f for water vapor are proportional to vapor density, but are approximately $\epsilon_i = 4.9$ and $\epsilon_f = 88$ for liquid water at relevant temperatures in the Earth's environment (Klein and Swift, 1977; Ulaby et al., 1986). The real and imaginary components of the electric susceptibility, $X_e(\omega) = \tilde{\epsilon}_r(\omega) - 1$, of liquid water are shown as a function of frequency in Figure 10. Note that ϵ_i represents the infinite frequency limit of the real part of the relative permittivity,



Radiation (Natural) Within the Earth's Environment,

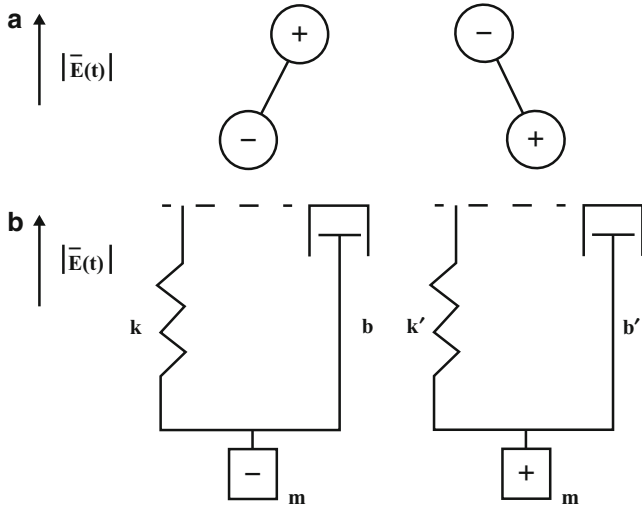
Figure 10 Debye relaxation model of the dielectric susceptibility of liquid water at 0 °C. The dielectric susceptibility at frequencies below ~1 GHz is consistent with all the water molecules following the orientation of the imposed electric field without significant loss. At frequencies above ~100 GHz, none of the water molecules can follow the imposed field, again, without significant loss. Maximum loss occurs at the maximum in the imaginary component of dielectric susceptibility. This maximum moves from 9 GHz at 0 °C to 17 GHz at 20 °C.

ϵ_f represents the zero frequency limit, and f_o is the frequency where $\epsilon'_r = \frac{\epsilon_f + \epsilon_i}{2}$.

The Lorentz process

Bending or stretching a bond between atoms also stores energy and, therefore, offers the possibility of a resonant process (Bohren and Huffman, 1983). For example, imagine the water molecule subjected to an oscillating electric field at a sufficiently high frequency that little rotation occurs. $\bar{E}(t)$ will cause water molecules to store elastic energy through bending and stretching of O–H bonds as shown schematically in Figure 11. Like rotation processes, specific vibrational resonances depend upon transitions specified by quantum mechanics, but the process can be understood without appealing to quantum mechanics through use of the model proposed by Lorentz in the early 1900s (e.g., Bohren and Huffman, 1983). The Lorentz model employs multiple spring–damper–mass systems, like those shown in Figure 11b. The systems are forced by an electric field, $\bar{E}(t)$, accelerating the charged masses. The dynamical equation for one system would be

$$m\ddot{z}(t) + b\dot{z}(t) + kz(t) = eE(t), \quad (44)$$



Radiation (Natural) Within the Earth's Environment,
Figure 11 Schematic of the Lorentz vibrational model of dielectric susceptibility. At frequencies well above those that permit a polyatomic molecule to align with an imposed electric field (a), the response of the molecule can be a vibrational mode modeled as a mass–spring–damper system forced by the imposed electric field (b). The system exchanges potential energy stored in the spring with kinetic energy stored in the motion of the mass. Unlike the Debye system, this system, if not overdamped, exhibits a resonance at angular frequency $\omega_o = \sqrt{\frac{k}{m}}$. Without damping, the system will attempt to store infinite energy at its resonant frequency.

where m is the atomic mass, b is a damping constant, k is a spring constant, and e is the electric charge. The number of dots over the coordinate, z , indicates the order of the time derivative. Polarization, $p(\omega)$, in the frequency domain, takes the form

$$\begin{aligned} p(\omega) &= e z(\omega) \\ &= \epsilon_o X_e(\omega) E(\omega) \end{aligned} \quad (45)$$

where $z(\omega)$ is the Fourier transform of $z(t)$ and $X_e(\omega)$ is the electric susceptibility in the frequency domain.

The dynamical equation in the frequency domain is

$$m(j\omega)^2 z(\omega) + b(j\omega) z(\omega) + k z(\omega) = e E(\omega), \quad (46)$$

and the frequency domain response to forcing becomes

$$p(\omega) = \frac{\left(\frac{e^2}{m}\right) E(\omega)}{(\omega_o^2 - \omega^2) + j\gamma\omega} \quad (47)$$

where damping coefficient, γ , replaces $\frac{b}{m}$ and resonant frequency, ω_o , replaces $\sqrt{\frac{k}{m}}$.

By recasting Equation 47 in the form of Equation 45 and writing $X_e(\omega)$ in polar form

$$X_e(\omega) = |X_e(\omega)| e^{-j\theta(\omega)}, \quad (48)$$

frequency-dependent amplitude, $X_e(\omega)$, and phase, $\theta(\omega)$, are

$$\begin{aligned} |X_e(\omega)| &= \frac{\left(\frac{e^2}{\epsilon_o m}\right)}{\sqrt{(\omega_o^2 - \omega^2)^2 + \gamma^2 \omega^2}} \\ \theta(\omega) &= \tan^{-1}\left(\frac{\gamma\omega}{\omega_o^2 - \omega^2}\right). \end{aligned} \quad (49)$$

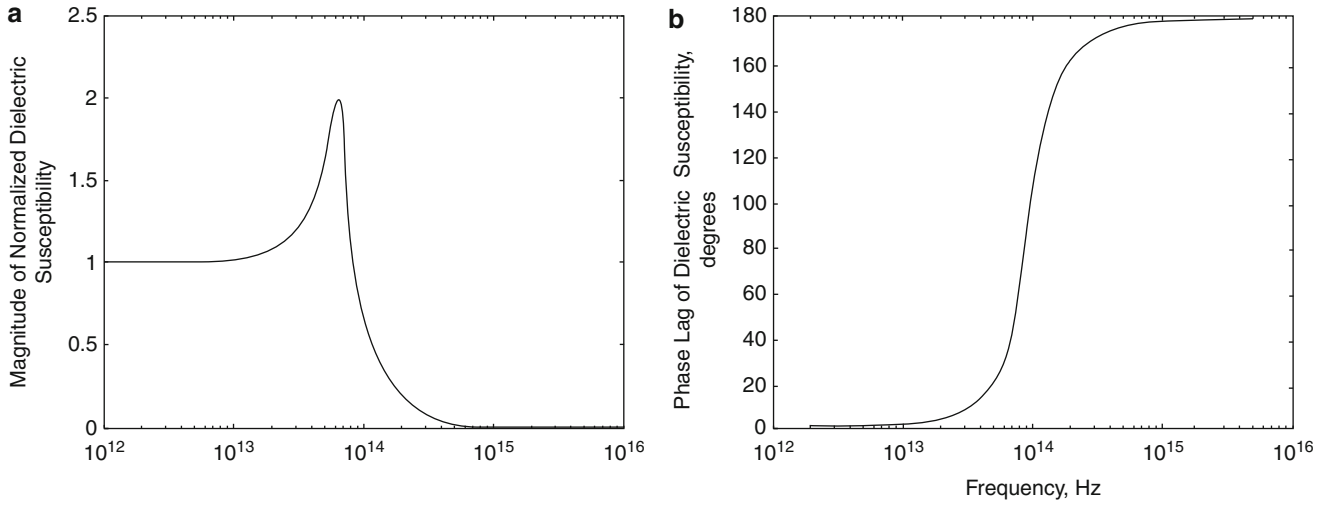
The dielectric susceptibility exhibited by this mass–spring–damper model, reflected in Equation 49, is illustrated in Figure 12 for a resonant frequency of 100 THz ($\omega_o = 2\pi \cdot 10^{14}$) and a damping coefficient, $\gamma = 0.5 \omega_o$. For $\omega \ll \omega_o$, the system responds purely as a spring with no energy dissipated and no phase lag. For $\omega = \omega_o$, the system exhibits maximum susceptibility, i.e., a resonance, most energy dissipated, and a phase lag of $\pi/2$. For $\omega \gg \omega_o$, the damping element dominates the system, the system cannot follow the forcing, there is no energy dissipated, and the phase lag approaches π .

The water vapor molecule exhibits reorientation processes, or Debye behavior, at 22.235 and 183.3 GHz (e.g., Waters, 1976). The O_2 molecule exhibits reorientation processes, or Debye behavior, as a series of features referred to as the 60 GHz complex and at 118.75 GHz (e.g., Rosenkranz, 1975). These Debye processes in water vapor and oxygen are the basis of most satellite remote sensing of atmospheric water vapor and temperature profiles. The CO_2 molecule exhibits a vibrational process, or Lorentz behavior, at 1.61 μm (186 THz) (e.g., Tolton and Plouffe, 2001). This Lorentz process in CO_2 was to be the basis of the first Orbiting Carbon Observatory (OCO), which experienced a launch failure in February 2009. The 1.62 μm band is well above the thermal infrared band, but the feature is included here because of the importance of atmospheric carbon sensing.

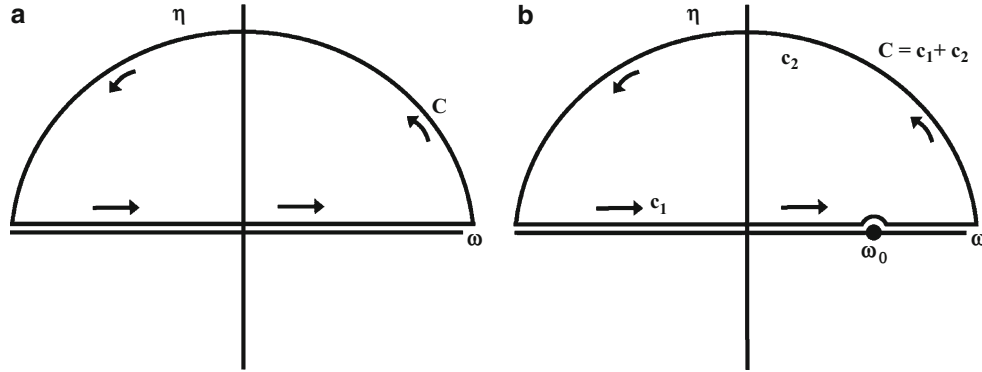
(v) Real and imaginary components of frequency domain EM constitutive properties are related through Kramers–Kronig relationships.

The Kramers–Kronig relationships represent constraints on the frequency domain behavior of a system that is real in the time domain and causal, i.e., system response cannot lead the causative impulse (e.g., Bohren and Huffman, 1983). Mathematical literature refers to such relationships as Poisson's formulas (e.g., Morse and Feshbach, 1953), and systems theory literature refers to them as Hilbert transform relations (e.g., Guillemin, 1949).

Development of the Kramers–Kronig relations for dielectric constitutive properties is eased if the problem can be cast in a form where the constitutive relation represents a difference between the response of a dielectric material and that of a vacuum. A useful consequence of this approach is that the difference property



Radiation (Natural) Within the Earth's Environment, Figure 12 Lorentz model of dielectric susceptibility. This example is based upon a resonant frequency of $f_o = 10^{14}$ Hz and a damping coefficient $\gamma = 0.5 \times 2\pi f_o$.



Radiation (Natural) Within the Earth's Environment, Figure 13 Cauchy integral paths. (a) shows a path for an infinite frequency time domain dielectric susceptibility of zero – essentially a self-consistent statement that the system cannot respond to an infinite frequency. (b) shows a path for dielectric susceptibility at frequency ω_o yielding the Kramers–Kronig relations.

goes to zero at infinite frequency because real materials cannot mechanically respond to an infinite frequency. Electric susceptibility in the frequency domain, $X_e(\omega) = \tilde{\epsilon}_r(\omega) - 1$, exhibits this behavior as illustrated in Figures 10 and 12 for both the Debye and Lorentz processes.

Electric susceptibility, the physical change in a system that results in electric polarization when an electric field is applied, is causal in that it is zero prior to an electric field impulse, i.e., in the time domain, $X_e(t - t_o) = X_e(\tau) = 0$ for an impulse at time $t \leq t_o$ or, equivalently, $\tau \leq 0$. It is mathematically convenient to define a complex frequency, $\tilde{\omega} = \omega - j\eta$, so that

$$e^{-j\tilde{\omega}\tau} = e^{-j\omega\tau} e^{-\eta\tau}, \quad (50)$$

where the first term to the right of the equal sign is a wave and the second term is an attenuation for $\eta > 0$.

The Fourier transform from time domain to complex frequency domain

$$X_e(\tilde{\omega}) = \int_0^\infty X_e(\tau) e^{-j\tilde{\omega}\tau} d\tau, \quad (51)$$

converges because $X_e(\omega)$ exists and $X_e(\tilde{\omega}) < X_e(\omega)$ for $\eta > 0$. That is, $X_e(\tilde{\omega}) = \tilde{\epsilon}_r(\tilde{\omega}) - 1$ is everywhere analytic (always finite) in the upper half of the ω and η plane shown in Figure 13a. The inverse transform of $X_e(\tilde{\omega})$ for $\tau = 0$ is

$$X_e(\tau = 0) = \int_C X_e(\tilde{\omega}) d\tilde{\omega} = 0 \quad (52)$$

by Cauchy's Theorem because contour, C , does not enclose a pole. Physically, this is equivalent to saying that the medium cannot respond instantaneously to an impulse

or, equivalently, that the medium cannot respond to an infinite frequency.

We introduce a pole in the susceptibility relation to explore the system response at frequency ω_o , i.e., $X'_e(\tilde{\omega}) = \frac{X_e(\tilde{\omega})}{\tilde{\omega} - \omega_o}$, which has a pole at $\tilde{\omega} = \omega_o$. The integral of $X'_e(\tilde{\omega})$ around the contour $C = C_1 + C_2$ in Figure 13b is the sum of the integrals along contours C_1 and C_2 . There are no poles within the contour, which means the integral around C is zero and the integral around C_2 is also zero because $X_e(\tilde{\omega}) \rightarrow 0$ as $\omega \rightarrow \infty$ and $\eta \rightarrow \infty$. That is, the integral along C_1 ($\eta = 0$) becomes

$$P \int_{\omega=-\infty}^{\omega=+\infty} X'_e(\omega) d\omega + j\pi X_e(\omega_o) = 0, \quad (53)$$

where P designates the Cauchy principal value and the second term represents the outcome of a Cauchy integral halfway around a pole at $\tilde{\omega} = \omega_o$. The conjugate symmetry of $X_e(\omega)$ (a result of its time domain version being real and causal) permits

$$\begin{aligned} P \int_{\omega=-\infty}^{\infty} X_e(\omega) \omega - \omega_o d\omega &= P \int_{-\infty}^{\infty} X_e(\omega) \omega - \omega_o d\omega \\ &+ P \int_0^{\infty} X_e(\omega) \omega - \omega_o d\omega \\ &= P \int_{-\infty}^0 X_e(-\omega') - (\omega' + \omega_o) d(-\omega') \\ &+ P \int_0^{\infty} X_e(\omega) \omega - \omega_o d\omega \\ &= \int_0^{\infty} -X_e^*(\omega') (\omega' + \omega_o) d(\omega') \\ &+ P \int_0^{\infty} X_e(\omega) \omega - \omega_o d\omega \end{aligned} \quad (54)$$

where $\omega' = -\omega$. The independent variables ω' and ω in the integrals on the right side of Equation 54 are dummy variables for the purposes of the integrals and can be combined to yield

$$P \int_{\omega=-\infty}^{\infty} \frac{X_e(\omega)}{\omega - \omega_o} d\omega = P \int_0^{\infty} \left\{ \frac{-X_e^*(\omega) (\omega - \omega_o) + X_e(\omega) (\omega + \omega_o)}{\omega^2 - \omega_o^2} \right\} d\omega. \quad (55)$$

Expanding the second term in Equation 53,

$$j\pi X_e(\omega_o) = j\pi(\varepsilon'_r(\omega_o) - 1) + j\pi(-j\varepsilon''_r(\omega_o)), \quad (56)$$

substituting the result of Equation 55 for the first term in Equation 53, and separating the real and imaginary parts of Equation 53 yield

$$\begin{aligned} \varepsilon'_r(\omega_o) &= 1 + \frac{2}{\pi} P \int_0^{\infty} \frac{\omega \varepsilon''_r(\omega)}{\omega^2 - \omega_o^2} d\omega \\ \varepsilon''_r(\omega_o) &= -\frac{2\omega_o}{\pi} P \int_0^{\infty} \frac{\varepsilon'_r(\omega) - 1}{\omega^2 - \omega_o^2} d\omega. \end{aligned} \quad (57)$$

Equations 57 are the Kramers–Kronig relations for relative permittivity. Equivalently, knowledge of either the real or the imaginary component of relative permittivity at all frequencies is equivalent to knowing the other component at any frequency, ω_o . In practice, it is usually adequate to know the first component over a frequency range of several decades centered upon ω_o to infer the other component. These relationships are often used to derive an unmeasured component from experimental knowledge of the other component. The reader is invited to verify that the Debye and Lorentz models satisfy the Kramers–Kronig relations.

(vi) Emissive power is a consequence of the imaginary component of relative permittivity.

The frequency-dependent absorption coefficient is a product of relationships among frequency domain electromagnetic fields. These relationships define the extinction of an electromagnetic wave and, consequently, the frequency-dependent absorption of power. The frequency domain expressions of Maxwell's Equations for charge-neutral media are

$$\begin{aligned} \nabla \times \bar{E}(\bar{r}, \omega) &= -j\omega \bar{B}(\bar{r}, \omega) \\ \nabla \times \bar{H}(\bar{r}, \omega) &= \bar{J}(\bar{r}, \omega) + j\omega \bar{D}(\bar{r}, \omega) \\ \nabla \cdot \bar{D}(\bar{r}, \omega) &= 0 \\ \nabla \cdot \bar{B}(\bar{r}, \omega) &= 0. \end{aligned} \quad (58)$$

where $\bar{E}(\bar{r}, \omega)$ and $\bar{H}(\bar{r}, \omega)$ are the electric and magnetic vector fields, $\bar{D}(\bar{r}, \omega)$ is the electric displacement vector field, $\bar{B}(\bar{r}, \omega)$ is the magnetic induction vector field, and $\bar{J}(\bar{r}, \omega)$ is the electric current density vector field. Unlike imaginary numbers in the time domain where their use often appears as a mathematical shortcut with limited physical meaning, imaginary numbers in the frequency domain have a precise physical meaning. For example, $-j$ in the Faraday Law (the first equation of Equation 58) denotes a 90° phase lead of the magnetic induction relative to the curl of the electric field, and $+j$ in the Ampere/Maxwell's Law (the second equation of Equation 58) denotes a 90° phase lag of the electric displacement relative to the curl of the magnetic field and the current density.

Materials found in significant abundances in the natural environment predominantly exhibit linear constitutive relationships at microwave frequencies, i.e.,

$$\begin{aligned} \bar{D}(\bar{r}, \omega) &= \varepsilon_o \bar{\varepsilon}_r(\bar{r}, \omega) \bar{E}(\bar{r}, \omega) \\ \bar{B}(\bar{r}, \omega) &= \mu_o \bar{H}(\bar{r}, \omega) \\ \bar{J}(\bar{r}, \omega) &= \sigma(\bar{r}) \bar{E}(\bar{r}, \omega). \end{aligned} \quad (59)$$

Magnetic permeability is vacuum permeability, μ_o , because natural media are predominantly nonmagnetic. The dominant electrical conduction process for natural media at temperatures found in the Earth's environment is either electron/hole migration at frequencies above 1 GHz or ion migration at frequencies well below 1 GHz. Interest in this section lies above 1 GHz where electron/hole migration dominates. At a given

temperature, these processes are approximately linear with the imposed electric field and independent of frequency over relevant frequencies so that conductivity, $\sigma(\bar{r})$, can be modeled as a real constant dependent only upon position vector, $\bar{r}$. The universality of this physical model for electric conductivity is unimportant in the context of this section because, for computational convenience, $\sigma(\bar{r})$ will be subsumed within a combined dielectric loss term where the contribution from electric conductivity is small relative to actual dielectric loss for the overwhelming majority of media of interest.

A combination of the frequency domain Maxwell's Equations and the vector identity, $\nabla \times \nabla \times \bar{E}(\bar{r}, \omega) = \nabla \nabla \cdot \bar{E}(\bar{r}, \omega) - \nabla^2 \bar{E}(\bar{r}, \omega)$, yields the Helmholtz wave equation in terms of the electric field:

$$(\nabla^2 + \omega^2 \mu_o \epsilon_o \tilde{\epsilon}_r(\omega)) \bar{E}(\bar{r}, \omega) = 0. \quad (60)$$

where electric conductivity has been subsumed into the definition of relative permittivity, i.e.,

$$\tilde{\epsilon}(\omega)_r = \epsilon'_r(\omega) - j \left(\epsilon''_r + \frac{\sigma}{\epsilon_o \omega} \right). \quad (61)$$

The conduction component, $\frac{\sigma}{\epsilon_o \omega}$, is typically much smaller than the dielectric component, ϵ''_r , at all thermally radiant frequencies. For simplicity, the conduction component is often ignored in further development of the loss term, a practice followed here, but users of these concepts should be aware that the imaginary component of relative permittivity potentially includes an electric conduction term that exhibits an inverse frequency dependence.

A plane wave traveling in the $\hat{k}$ direction can be written

$$\bar{E} = \hat{e} E_o e^{-j \hat{k} \cdot \bar{r}}. \quad (62)$$

The electric field direction is designated by unit vector $\hat{e}$ normal to $\hat{k}$, and the wave vector is $\bar{k} = \hat{k} k$, where k is the wave number (or angular spatial frequency), and the explicit dependences upon position, $\bar{r}$, and frequency, ω , have been suppressed. In a Cartesian frame, $\hat{x}$, $\hat{y}$, and $\hat{z}$, the phase term of the electric field becomes

$$\begin{aligned} \bar{k} \cdot \bar{r} &= (\hat{x} k_x + \hat{y} k_y + \hat{z} k_z) \cdot (\hat{x} x + \hat{y} y + \hat{z} z) \\ &= k_x x + k_y y + k_z z \end{aligned} \quad (63)$$

where k_x , k_y , and k_z are the orthogonal components of $\mathbf{k}$. The Laplacian of the electric field is

$$\begin{aligned} \nabla^2 \bar{E} &= \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \bar{E} \\ &= - \left(k_x^2 + k_y^2 + k_z^2 \right) \bar{E} \\ &= -k^2 \bar{E}. \end{aligned} \quad (64)$$

Substituting Equation 64 into the Helmholtz wave equation (Equation 60) yields

$$(-k^2 + \omega^2 \mu_o \epsilon_o \tilde{\epsilon}_r) \bar{E} = 0. \quad (65)$$

The characteristic equation of Equation 65, the dispersion relation, becomes

$$\begin{aligned} k &= \omega \sqrt{\mu_o \epsilon_o} \sqrt{\tilde{\epsilon}_r} \\ &= \frac{\omega}{c} \sqrt{\tilde{\epsilon}_r} \\ &= k_o \tilde{n} \end{aligned} \quad (66)$$

where $\sqrt{\mu_o \epsilon_o}$ is the inverse of the vacuum speed of light, c ; $\frac{\omega}{c}$ is the vacuum wave number, k_o ; and $\tilde{n}$, the complex index of refraction, is the square root of the relative permittivity. The complex index of refraction, $\tilde{n}$, is defined $\tilde{n} = n - j\kappa$ where n is the familiar index of refraction and κ is the wave extinction coefficient. The role of the complex index of refraction, $\tilde{n}$, in wave propagation is illustrated as follows:

$$\begin{aligned} \bar{E} &= \hat{e} E_o e^{-j \hat{k} \cdot \bar{r}} \\ &= \hat{e} E_o e^{-j (\hat{k} \cdot \bar{r}) k r} \\ &= \hat{e} E_o e^{-j (\hat{k} \cdot \bar{r}) k_o \tilde{n} r} \\ &= \hat{e} E_o e^{-j (\hat{k} \cdot \bar{r}) k_o n r} e^{-(\hat{k} \cdot \bar{r}) k_o \kappa r}. \end{aligned} \quad (67)$$

The first exponential term in the last line of Equation 67 is phase, and the second exponential term is attenuation. If the relative permittivity, $\tilde{\epsilon}_r$, is real, its square root, $\tilde{n}$, is real. That is, $\kappa = 0$, i.e., there would be no extinction of the wave.

Converting from propagation of an electric field to radiant power requires knowledge of the magnetic field. Combining the Faraday and the Ampere/Maxwell laws yields

$$\begin{aligned} \bar{H} &= \hat{h} \sqrt{\frac{\epsilon_o}{\mu_o}} \tilde{n} E_o e^{-j k_o \tilde{n} r} \\ &= \hat{h} \frac{\tilde{n} E_o}{\eta_o} e^{-j k_o \tilde{n} r} \end{aligned} \quad (68)$$

where $\hat{h}$ is a unit vector indicating the direction of the magnetic field and $\eta_o = \sqrt{\frac{\mu_o}{\epsilon_o}}$ is the vacuum impedance to an electromagnetic wave. It is easily shown that $\hat{h}$ is mutually perpendicular to $\hat{k}$ and $\hat{e}$. Average radiant power, expressed as a Poynting vector, is half the real part of the product of the electric field crossed with the complex conjugate of the magnetic field, i.e.,

$$\begin{aligned} \bar{S} &= \frac{1}{2} \text{Re} \{ \bar{E} \times \bar{H}^* \} \\ &= \frac{1}{2} \hat{k} \text{Re} \{ E H^* \} \\ &= \frac{1}{2} \hat{k} \text{Re} \left\{ E \left(\frac{\tilde{n}}{\eta_o} E \right)^* \right\} \\ &= \frac{1}{2} \hat{k} \text{Re} \left\{ E E^* \frac{\tilde{n}^*}{\eta_o} \right\} \\ &= \hat{k} \frac{E_o^2}{2 \eta_o} n e^{-2 k_o \kappa r}. \end{aligned} \quad (69)$$

The exponential term in the last line of Equation 69 represents absorption of the radiant power. Defining the absorption coefficient, $\kappa_{fa} = 2 k_o \kappa$, yields the familiar expression

$$\bar{S} = \hat{k} S_o e^{-\kappa_{fa} r}, \quad (70)$$

where S_o is the power density at $r = 0$. Aspects of Equation 70 to remember are that the absorption coefficient, κ_{fa} , is a frequency domain concept and that it is nonzero only if either the dielectric contribution or the conduction contribution to the imaginary component of the relative permittivity is nonzero.

Scattering loss, often referred to as scatter darkening, introduces an additional multiplier to Equation 70 of the form $e^{-\kappa_{fs} r}$, where κ_{fs} is a frequency-dependent scattering coefficient. The sum of absorption and scattering coefficients is the frequency-dependent power extinction coefficient, $\kappa_{fe} = \kappa_{fa} + \kappa_{fs}$. Scattering is beyond the scope of this section, but it is appropriate to note that scattering influences emissive power only to the extent that scattering warms the emitting medium. Scattering effectively increases the length of ray paths within the scattering medium. Longer ray paths are subject to increased absorption so that the radiant brightness escaping from the scattering medium will be diminished. This scatter darkening contributes to the thermal temperature of the medium only for scattered energy that is a significant fraction of the total thermal energy of the scattering medium. In practice, this means radiant energy from the Sun or thermal infrared energy from sources within the Earth's environment. Scattering at millimeter and microwave wavelengths does not appreciably increase the thermal temperature of the medium and, therefore, does not influence the emissive power of the medium.

Summary

Natural radiation originating within the Earth's environment at frequencies between 1 GHz and 40 THz ($\sim 8 \mu\text{m}$, the upper end of the TIR band) is a consequence of Planck thermal emission. This radiation can be used to infer thermal and compositional states of the natural emitter and the media traversed by the ray after it leaves the natural emitter. A condition for Planck emission is local thermal equilibrium. Emissive power equaling absorption at every frequency, i.e., the principle of detailed balance (e.g., McQuarrie, 2000), is a condition of thermal equilibrium. A corollary of that principle is that, in the absence of absorption and scattering, radiant brightness temperature is invariant along a ray. Emissive power, an intrinsic property of a medium, is determined by the medium's dielectric loss, conduction loss, and thermal temperature. Dielectric and electric conduction constitutive properties are most intuitive when expressed in the frequency domain. The real and imaginary components of frequency domain constitutive

properties carry physical meaning and are related through Kramers–Kronig relationships. The contribution of the constitutive properties to emissive power originates in the imaginary component of relative permittivity.

Bibliography

- Bohren, C. F., and Huffman, D. R., 1983. *Absorption and Scattering of Light by Small Particles*. New York: Wiley.
- Chandrasekhar, S., 1960. *Radiative Transfer*. New York: Dover.
- Debye, P., 1929. Der Lichtdruck auf Kugeln von beliebigem material. *Annals of Physics*, **30**, 57–136.
- Grant, E., Buchanan, T., and Cook, H., 1957. Dielectric behavior of water at microwave frequencies. *The Journal of Chemical Physics*, **26**, 156–161.
- Guillemin, E. A., 1949. *The Mathematics of Circuit Analysis*. New York: Wiley.
- Klein, L. A., and Swift, C. T., 1977. An improved model for the dielectric constant of sea water at microwave frequencies. *IEEE Transactions on Antennas and Propagation*, **AP-25**, 104–111.
- Kraus, J. D., 1966. *Radio Astronomy*. New York: McGraw-Hill.
- Liebe, H. J., 1969. Calculated tropospheric dispersion and absorption due to the 22-GHz water vapor line. *IEEE Transactions on Antennas and Propagation*, **AP-17**, 621–627.
- McQuarrie, D. A., 2000. *Statistical Mechanics*. Mill Valley: University Science Books.
- Modest, M. F., 2003. *Radiative Heat Transfer*. San Diego: Academic.
- Morse, P. M., and Feshbach, H., 1953. *Methods of Theoretical Physics*. New York: McGraw-Hill.
- Planck, M., 1909. Zur Theorie der Wärmestrahlung. *Annalen der Physik*, **336**, 758–768.
- Rosenkranz, P. W., 1975. Shape of the 5 mm oxygen band in the atmosphere. *IEEE Transactions on Antennas and Propagation*, **AP-23**, 498–506.
- Stogryn, A., 1971. Equations for calculating the dielectric constant of saline water. *IEEE Transactions on Microwave Theory and Techniques*, **MIT-19**, 733–736.
- Tolton, B. T., and Plouffe, D., 2001. Sensitivity of radiometric measurements of the atmospheric CO₂ column from space. *Applied Optics*, **40**, 1305–1313.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing: Active and Passive*. Reading: Addison-Wesley. Microwave Remote Sensing Fundamentals and Radiometry, Vol. 1.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1986. *Microwave Remote Sensing: Active and Passive*. Norwood: Artech House. From Theory to Applications, Vol. III.
- Van de Hulst, H. C., 1981. *Light Scattering by Small Particles*, 2nd edn. Mineola: Dover.
- Waters, J. W., 1976. Absorption and emission of microwave radiation by atmospheric gases. In Meeks, M. L. (ed.), *Methods of Experimental Physics Part B*. New York: Academic. Academic. Radio Astronomy, Sec 2.3, Vol. 1–12.

Cross-references

[Radiation Sources \(Natural\) and Characteristics](#)
[Radiation, Galactic, and Cosmic Background](#)
[Radiation, Solar and Lunar](#)

RADIATION SOURCES (NATURAL) AND CHARACTERISTICS

Anthony England
College of Engineering, University of Michigan,
Ann Arbor, MI, USA

Synonyms

Brightness; Radiance

Definition

Natural radiation is an electromagnetic product of physical processes in the Earth or space environment and, as such, is often a remotely sensed indicator of aspects of those processes.

Natural radiation sources, characteristics

Remote sensing techniques that sense electromagnetic radiation are referred to as “active” if they employ artificial sources or “passive” if they employ natural sources. **RADAR** and **LiDAR** are examples of active techniques. Sources for “passive” techniques, that is, those employing natural radiation, are classified as thermal and nonthermal. Thermal sources include Planck emission from any dielectric medium whose temperature is above absolute zero (Planck, 1909) and ionized hydrogen emission from hot plasmas like those found at the surfaces of the Sun and stars (e.g., Hjellming, 1974). Nonthermal sources are of little importance in remote sensing of the Earth except for studies using radio emission from lightning (sferics) or from the ionosphere (synchrotron radiation). Nonthermal sources are significant in radio astronomy where synchrotron radiation, a result of interactions between plasmas and magnetic fields, can be strong and diagnostic of stellar or galactic processes (Salter and Brown, 1974; Harwit, 1988).

Planck emission is leakage of radiant energy from a dielectric whose size is large with respect to a wavelength, whose temperature is uniform over distances that are large with respect to a wavelength, and whose internal radiant energy is in thermal equilibrium. The Planck brightness for polarization, p , within a homogeneous dielectric medium, readily derived using statistical mechanics (e.g., McQuarrie, 2000), is given by

$$B_p(f, T, n) = \left(\frac{n}{c}\right)^2 \frac{h f^3}{(e^{\frac{hf}{kT}} - 1)} \text{Wm}^{-2}\text{Hz}^{-1}\text{sr}^{-1}, \quad (1)$$

where f is frequency (Hz), T is temperature (K), n is index of refraction, c is vacuum speed of light (2.9979×10^8 m/s), h is Planck's constant (6.6256×10^{-34} J · s), and k is Boltzmann's constant (1.3805×10^{-23} J/K).

If a ray can leak out of the dielectric medium without reflection at the dielectric-vacuum interface, then the medium is classified as a blackbody and its brightness is given by

$$B_p^b(f, T) = B_p(f, T, n = 1) \text{Wm}^{-2}\text{Hz}^{-1}\text{sr}^{-1}, \quad (2)$$

where superscript b identifies blackbody radiance. Even though blackbody brightness is independent of polarization, subscript p reminds us that B_p^b refers to single-polarization brightness and that the microwave brightness of a non-black (or “gray”) surface is typically polarization dependent. Total blackbody spectral brightness would be $2B_p^b$.

Figure 1 shows blackbody brightness spectra, $B_p^b(f, T)$, plotted against wavelength using the relationship, $\lambda = c/f$, for dielectrics whose temperatures are 300 K (approximately the physical temperature of the Earth or Moon) and 6,000 K (approximately the physical temperature of the visible disk, or photosphere, of the Sun). The ionized hydrogen brightness of the solar photosphere looks nearly like a 6,000 K Planck emitter for UV through millimeter wavelengths. The exceptions are several hundred dark lines, or Fraunhofer lines, caused by absorption in the cooler gaseous shell, or chromosphere, surrounding the Sun. These lines are too narrow to be factors in UV, visible, or reflectance infrared (near-IR) sensing of the Earth and are too narrow to be depicted at the scale of the 6,000 K spectrum in Figure 1.

The solar brightness at wavelengths longer than a millimeter increases and varies spectacularly with the 11 year solar cycle (limits depicted in Figure 1). At solar minimum, spectral regions of the microwave brightness reach the equivalent of a 10^6 K Planck emitter. At solar maximum, the peak brightness in a limited band can exceed the equivalent of a 10^{10} K emitter, a four orders of magnitude difference (Kraus, 1966).

There are three salient characteristics of Planck, or “blackbody,” emission that are fundamental to remote sensing (e.g., Ulaby et al., 1981). The first salient characteristic, called Wien's displacement law, is the maximum brightness of a Planck spectrum that occurs at a wavelength that is inversely proportional to the emitter's thermal temperature, that is,

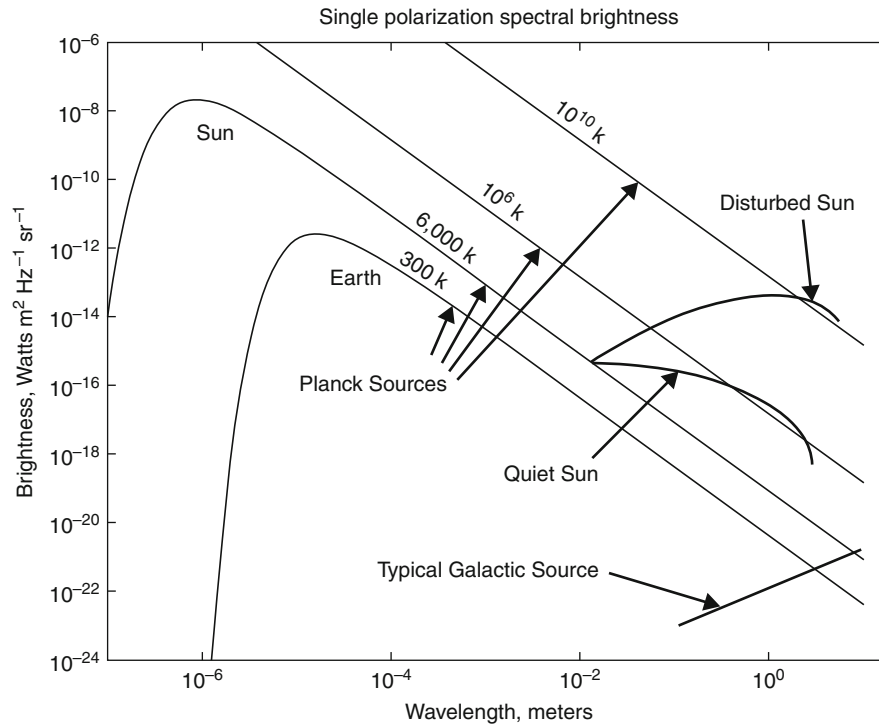
$$\lambda_{\max} = \frac{C_w}{T} \text{ m}, \quad (3)$$

where Wien's displacement constant, C_w , in vacuum is 2.899×10^{-3} m K. Given that most temperatures in the Earth's environment are near 300 K, convention has defined thermal infrared (TIR) remote sensing of the Earth as centered about $\lambda_{\max} = 10 \mu$.

The second salient characteristic of Planck emission, called the Stefan-Boltzmann law, is the total spectral brightness integrated over all frequencies, and both polarizations are proportional to the fourth power of the thermal temperature, that is,

$$B_{\text{tot}} = \frac{\sigma T^4}{\pi} \text{ W} \cdot \text{m}^{-2} \cdot \text{sr}^{-1}, \quad (4)$$

where the Stefan-Boltzmann constant, σ , in vacuum is $5.673 \times 10^{-8} \text{ W} \cdot \text{m}^{-2} \cdot \text{K}^{-4} \cdot \text{sr}^{-1}$. Because the brightness of a “black” emitter is uniform in all directions, B_{tot} is



Radiation Sources (Natural) and Characteristics, Figure 1 Examples of natural source, single-polarization brightness spectra. The Earth and Moon exhibit a 300 K Planck radiance, which is typical of the Earth and Moon. Because of nonthermal processes in the solar corona, the solar spectrum deviates from a 6,000 K Planck curve at wavelengths longer than a centimeter. The brightness spectra of Planck emitters exhibit -2 slopes at longer wavelengths on a log-log plot. Nonthermal emitters, like the solar corona and galactic sources, typically exhibit other than -2 slopes.

a constant in the integral defining total radiant power lost through a square meter of a “black” surface.

$$\begin{aligned} \text{Radiant Power Lost} &= \int_{\phi=0}^{\phi=2\pi} \int_{\theta=0}^{\theta=\pi/2} (\sin \theta \, d\phi) \\ &\quad \times (B_{\text{tot}} \cos \theta) \, d\theta = \sigma T^4 \quad \text{W/m}^{-2}. \end{aligned} \quad (5)$$

This version of the Stefan-Boltzmann law is a staple of thermal design of engineered systems (e.g., Modest, 2003).

The third salient characteristic of Planck emission, called the Raleigh-Jeans approximation for any wavelength that is long with respect to λ_{max} , defines that single-polarization brightness is proportional to temperature. The relationship in vacuum is given by

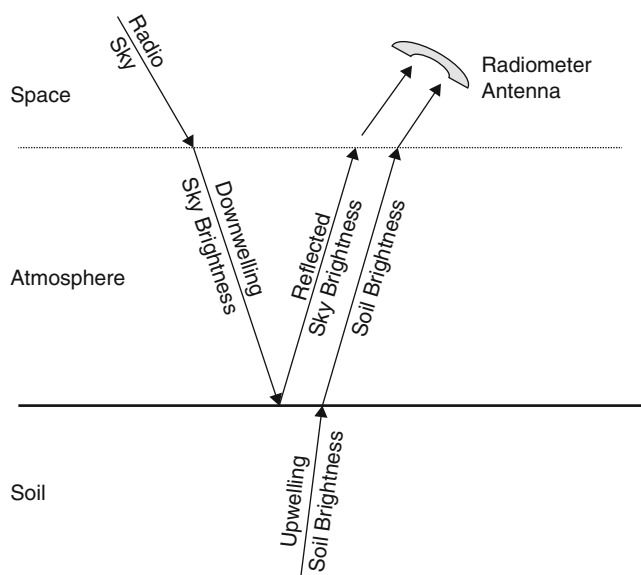
$$B_p(f, T) = \frac{kT}{\lambda^2} \text{ W} \cdot \text{m}^{-2} \cdot \text{Hz}^{-1} \cdot \text{sr}^{-1}. \quad (6)$$

Equation 6 suggests that a -2 slope appearing in a log-log plot of spectral brightness versus wavelength could be diagnostic of a Planck emitter. The distinction is exhibited in the differences between the observed spectral brightness of the Moon and Mars and those of the examples of galactic sources in Figure 1. The Moon and Mars, thermal emitters, exhibit slopes of -2 , while the galactic sources,

nonthermal emitters, exhibit positive spectral slopes. A -2 slope is also characteristic of the microwave brightness spectra of media in the Earth’s environment. Exceptions to the rule occur when volume scattering, surface scattering, or interference within relatively transparent layers modifies the Planck spectrum of dielectric media. These situations will be discussed more fully in later sections.

The natural radiation itself can be a diagnostic feature of an emitter. For example, the Planck-like spectrum defines the temperature at the surface of the solar optical disk as 6,000 K, and the Fraunhofer absorption lines in that spectrum are used to infer the composition of gases in the solar chromosphere (e.g., Ross and Aller, 1976). The microwave spectral brightness of the Earth’s atmosphere is used to infer temperature and moisture profiles in the atmosphere, and the brightness of soil at wavelengths longer than 10 cm is used to infer an integrated liquid moisture content of soils to depths of something like half a wavelength in the soil (e.g., Ulaby et al., 1986).

Natural radiation is also used to illuminate remotely sensed objects of interest. For example, solar brightness is the source of visible and near-infrared (reflectance IR) sensing in the Earth’s environment (e.g., Deering, 1989). There is also background illumination, like that depicted in the cartoon of retrieval of soil moisture (Figure 2). The microwave signature of soil moisture is primarily



Radiation, Electromagnetic, Figure 2 Cartoon of scheme for sensing soil moisture. The radio sky includes the ubiquitous 2.7 K cosmic background radiation plus a nonthermal contribution from galactic sources. The galactic contribution is only significant at frequencies near and below 1 GHz. The galactic contribution is greatest from the center of the Milky Way and least from its poles. At 1.4 GHz, the maximum contribution is about 10 K and the minimum is below the cosmic background (Tiuri, 1966). Absorption and emission by the atmosphere is frequency dependent but relatively insignificant at the preferred soil moisture sensing frequency of 1.4 GHz (Ulaby et al., 1986). Emission from wet soil has a wetness-dependent dynamic range of tens of Kelvins at 1.4 GHz even through significant agricultural canopies (e.g., Hornbuckle and England, 2004).

due to Planck emission from wet soil, but satellite radiometers will sense Planck emission from the soil, sky brightness reflected by the soil, and modification of both by absorption and emission in the atmosphere between the soil and the radiometer. An accurate retrieval of soil moisture requires that all of these processes be considered. Such retrievals will be explored in later sections.

Conclusion

Natural radiation is the source for “passive” remote sensing techniques based upon electromagnetic radiation. These passive techniques are typically labeled gamma ray, X-ray, reflectance ultraviolet, reflectance visible, reflectance infrared, millimeter radiometry, and microwave radiometry. Techniques in the UV through microwave bands are significant for Earth remote sensing.

Natural radiation sources are either thermal or nonthermal. Thermal means that the radiant energy is derived from thermal processes in the emitting medium. Nonthermal usually means that the radiant energy is derived from an interaction between a plasma and a magnetic field, that is, synchrotron radiation. Thermal sources are

overwhelmingly the dominant source for remote sensing in the Earth’s environment.

The spectral brightness of a thermal source can be diagnostic of the source, for example, the temperature and salinity of standing water, or of the medium the rays pass through in route to the radiometer, for example, the temperature and water vapor profiles of the atmosphere.

Bibliography

- Deering, D. W., 1989. Field measurements of bidirectional reflectance. In Asrar, G. (ed.), *Theory and Applications of Optical Remote Sensing*. New York: Wiley.
- Harwit, M., 1988. *Astrophysical Concepts*, 2nd edn. New York: Springer.
- Hjellming, R. M., 1974. Radio stars. In Verschuur, G. L., and Kellermann, K. I. (eds.), *Galactic and Extragalactic Radio Astronomy*, 2nd edn. Berlin: Springer, pp. 381–438.
- Hornbuckle, B. K., and England, A. W., 2004. Radiometric sensitivity to soil moisture at 1.4 GHz through a corn crop at maximum biomass. *Water Resources Research*, **40**, W10204, doi:10.1029/2003WR002931.
- Kraus, J. D., 1966. *Radio Astronomy*. New York: McGraw-Hill.
- McQuarrie, D. A., 2000. *Statistical Mechanics*. Sausalito, CA: University Science.
- Modest, M. F., 2003. *Radiative Heat Transfer*. San Diego, CA: Academic.
- Planck, M., 1909. Zur Theorie der Wärmestrahlung. *Annalen der Physik*, **336**, 758–768.
- Ross, E. R., and Aller, L. H., 1976. The chemical composition of the sun. *Science*, **191**, 1223–1229.
- Salter, C. J., and Brown, R. L., 1974. Galactic nonthermal continuum emission. In Verschuur, G. L., and Kellermann, K. I. (eds.), *Galactic and Extragalactic Radio Astronomy*, 2nd edn. Berlin: Springer, pp. 1–36.
- Tiuri, M. E., 1966. Radio-telescope receivers. In Kraus, J. D. (ed.), *Radio Astronomy*. New York: McGraw-Hill.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing: Active and Passive*. Reading, MA: Addison-Wesley. Microwave Remote Sensing Fundamentals and Radiometry, Vol. I.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1986. *Microwave Remote Sensing: Active and Passive*. Norwood, CA: Artech House. From Theory to Applications, Vol. III.

Cross-references

[Radiation, Galactic, and Cosmic Background](#)
[Radiation \(Natural\) Within the Earth’s Environment](#)
[Radiation, Solar and Lunar](#)

RADIATION, ELECTROMAGNETIC

Frank S. Marzano
 Department of Information Engineering, Sapienza
 University of Rome, Rome, Italy
 Centre of Excellence CETEMPS, University of L’Aquila,
 L’Aquila, Italy

Definition

Electromagnetic radiation is the means by which *information* is transmitted from an object to the sensor

through an emitted, transmitted, or reflected electric and/or magnetic field. The *interaction* of the electromagnetic waves with the environment is strongly dependent on the wavelength and the geometrical, electrical, magnetic, and conductive properties of the interacting medium. Electromagnetic waves in different frequency bands tend to excite different interaction mechanisms such as electronic, molecular, or conductive. Electromagnetic radiation is the most commonly used means to retrieve information about an object by means of remote sensing techniques.

Introduction

Remote sensing is the methodology to remotely retrieve information about an object (Ulaby et al., 1981). This information is acquired by measuring changes that the object under investigation imposes on the interacting field. The latter can be an electromagnetic, acoustic, or gravitational field. The electromagnetic methods, which are the most commonly used, cover the whole electromagnetic spectrum from radio frequency waves to gamma rays through microwaves, submillimeter waves, and far-infrared, near-infrared, visible, ultraviolet, and X-ray waves (Elachi, 1987).

Electromagnetic radiation spectrum

From the classic electromagnetic theory point of view, the *electromagnetic radiation* at a given frequency is a wave phenomenon due to the contemporary propagation of periodic perturbations of an electric and a magnetic field, in free space oscillating in planes perpendicular to each other (Ishimaru, 1991). For a wave periodic phenomenon, it is usual to introduce the wavelength λ and the frequency ν so that the wave velocity c in free space is expressed by

$$c = \frac{\lambda}{T} = \lambda\nu = \lambda \frac{\omega}{2\pi} = \frac{\omega}{\beta} \quad (1)$$

where T is the time period, ω is angular pulsation, and β the propagation constant (or wave number). From Eq. 1, an electromagnetic wave with a higher frequency will exhibit a shorter wavelength. The *electromagnetic spectrum* is divided into a number of spectral regions, each of them conventionally labeled as in Table 1 (Ulaby et al., 1981). The latter emphasizes the spectral bands of interest for remote sensing purposes.

The *radio wave region* covers wavelengths shorter than 10 cm and is used by active radio sensors such as imaging radars, altimeters, and sounders. The main mechanism of interaction is due to scattering, reflection, conduction, and ionospheric effects.

The *microwave region* is governed by molecular rotation, particularly at shorter wavelength, and the main interaction mechanisms include thermal emission, scattering, and conduction. The major applications refer to microwave radars, radiometers, and spectrometers. It is customary to subdivide the microwave region in

Radiation, Electromagnetic, Table 1 Electromagnetic spectrum and its nomenclature

Band	Frequency range (ν)	Wavelength range (λ)
Radio wave	<3 GHz	>10 cm
Microwave	3–300 GHz	1 mm–10 cm
Submillimeter	300 GHz–3 THz	10 cm–0.1 mm
Infrared	3–428 THz	0.1 mm–700 nm
Visible	428–749 THz	700–400 nm
Ultraviolet	749 THz–30 PHz	400–10 nm
X-rays	30 PHz–300 EHHz	10 nm–1 pm
Gamma rays	>300 EHHz	<1 pm

frequency bands, labeled by a letter: P (0.5–1 GHz), S (2–4 GHz), C (4–8 GHz), X (8.124 GHz), Ku (12.4–18 GHz), K (18–26.5 GHz), Ka (26.5–40 GHz), V (40–75 GHz), and W (75–90 GHz).

The *submillimeter and infrared region* covers a wavelength interval between 1 mm and 0.7 μ m. The infrared region is usually subdivided into far-infrared (15–1,000 μ m), thermal infrared (8–15 μ m), mid-infrared (3–8 μ m), and near-infrared (0.75–3 μ m). In this region, molecular rotation and vibration play an important role together with thermal emission. Imagers, spectrometers, radiometers, polarimeters, interferometers, and lasers are typical sensors exploiting this frequency region for remote sensing.

The *visible region* covers the electromagnetic spectrum portion visible by human eyes between 380 and 750 nm. This interval is usually subdivided into colors: violet (380–450 nm), blue (450–495 nm), green (495–570 nm), yellow (570–590 nm), orange (590–620 nm), and red (620–759 nm). In this region, electronic energy levels start to play a key role together with vibrational processes in the wave–matter interactions. Imagers, spectrometers, radiometers, polarimeters, and lasers are among the sensors operating in this region.

The *ultraviolet region* is characterized by interactions dominated by electronic processes. Major applications are related to planetary atmosphere sounding because of the string gas opacity at these wavelengths.

The *X-ray and gamma-ray regions* are related to wave–matter interaction involving atomic processes. Where the X photons are produced by energetic electrons, the gamma ones are generated by transitions within an atomic nucleus. The strong atmospheric opacity limits their use to planetary surfaces without atmosphere.

Electromagnetic radiation theory

Following classical physics, the macroscopic behavior of electromagnetic waves is governed by Maxwell's differential equations, which relate the electric field vector $\mathbf{E}(\mathbf{r},t)$ [V/m] to the magnetic field vector $\mathbf{H}(\mathbf{r},t)$ [A/m] in a given instant t and point, indicated by a position vector $\mathbf{r}$ (Tsang et al., 1985). In a linear, isotropic, homogeneous,

non-ferromagnetic, and dissipative medium, the Maxwell equations can be written as

$$\begin{cases} \nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \\ \nabla \times \mathbf{H} = -\varepsilon \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} + \mathbf{J}_i \\ \nabla \cdot \mathbf{E} = \frac{\rho_i}{\varepsilon} \end{cases} \quad (2)$$

where ∇ is the *nabla* spatial differential operator, μ [H/m] is the medium magnetic permeability, ε [F/m] is the medium electric permittivity, and σ is the medium electronic conductivity. The density current $\mathbf{J}_i$ and electric charge concentration ρ_i [C/m³] represent the impressed quantities, either generated by an electromagnetic source or known from proper boundary conditions. The last equation in Eq. 2 is the Gauss law. Previous equations state that a wave motion exists in the electric and magnetic force fields. In any region where there is a temporal change of the electric field, a magnetic field appears in that region as a conjugal partner and vice versa.

Free-space radiation

In absence of impressed quantities (electromagnetic sources) and in a non-dissipative medium (i.e., $\sigma = 0$), the previous Maxwell equations can be combined to derive the following *wave equation* for the electric field:

$$\nabla^2 \mathbf{E} - \mu \varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \quad (3)$$

where ∇^2 is the *Laplacian* spatial second-order differential operator. A dual wave equation for the magnetic field can be derived from Eq. 3, substituting $\mathbf{E}$ with $\mathbf{H}$. In case of cosinusoidal time-periodic waves and in an unbounded medium, a relevant solution of the previous wave equation is represented by the uniform plane wave expressed by

$$\mathbf{E}(\mathbf{r}, t) = \text{Re}[\mathbf{E}_0 e^{j\omega t - j\beta \cdot \mathbf{r}}] = \eta[\hat{\mathbf{r}} \times \mathbf{H}(\mathbf{r}, t)] \quad (4)$$

where $\mathbf{E}_0$ is the complex polarization state vector, $j = \sqrt{-1}$ is the imaginary unit, $\eta = \sqrt{\mu/\varepsilon}$ [Ω] is the *characteristic impedance* of the medium, and $\hat{\mathbf{r}}$ is the unit vector of the position vector $\mathbf{r}$. In Eq. 4, Re is the *real-part* operator, and the quantity $\Psi(\mathbf{r}, \omega) = \omega t - \beta \cdot \mathbf{r}$ is also called the plane-wave *phase function* with $\omega = 2\pi\nu$ the angular pulsation, proportional to the frequency ν , and $\beta = \beta \hat{\mathbf{r}}$ the propagation vector. The uniform plane wave expresses the typical characteristic of a transverse electromagnetic (TEM) wave where $\mathbf{E}$ and $\mathbf{H}$ are proportional, mutually orthogonal, and oscillate in a plane perpendicular to direction of propagation indicated by the propagation vector β . It is worth noting that the plane wave is a nonphysical solution due to its infinite energy; it is, however, important to express wave propagation properties (Ulaby et al., 1981).

In presence of electromagnetic sources and in a non-dissipative medium, the Maxwell equations in an

unbounded free space can be formally solved, and the monochromatic electric field $\mathbf{E}$ in spherical coordinates $\mathbf{r} = [r, \theta, \varphi]$ has the following expression:

$$\begin{aligned} \mathbf{E}(\mathbf{r}, t) &= \text{Re} \left\{ e^{j\omega t} \mathbf{F}_E(\theta, \varphi) \frac{e^{-j\beta r}}{r} \right\} \\ &= \eta[\hat{\mathbf{r}} \times \mathbf{H}(\mathbf{r}, t)] \end{aligned} \quad (5)$$

where $\mathbf{F}_E$ is the electric radiation vector function, indicating the polarization property and relating the radiated field to the electromagnetic sources in terms of a transform. The latter, for example, in case the sources are expressed in Cartesian coordinates, is explicitly a Fourier three-dimensional transform. The quantity $\Psi(\mathbf{r}, \omega) = \omega t - \beta r$ is the spherical-wave phase function in Eq. 5. The fact that electric radiation vector function $\mathbf{F}_E(\theta, \varphi)$ is not depending on the distance r is true only in the so-called far-field radiation zone (or Fraunhofer's zone) which is a region where r is larger than D_s^2/λ , with D_s the maximum size of the electromagnetic source. Equation 5 highlights the major properties of the radiated field in a free space: (1) it is a spherical wave; (2) its amplitude is inversely proportional to the distance r ; (3) it can be locally modeled as a TEM plane wave (Fung, 1994).

When the electromagnetic problem is solved either in a bounded medium or in presence of other sources and obstacles, the free-space solution is no more valid (Ishimaru, 1991). Proper initial and boundary conditions, describing the tangential components of the electromagnetic field on the boundaries and their initial time state, need to be imposed to the Maxwell equations. These conditions ensure the uniqueness of the electromagnetic solution within the considered volume.

Quantum radiation

Maxwell's theory fails to account for certain significant phenomena when the wave interacts with matter, especially at very short wavelengths where electronic, atomic, and molecular forces play a dominant role. In these cases, *quantum electrodynamics* is a theory more appropriate than classical electromagnetics to describe electromagnetic radiation (Elachi, 1988).

The electromagnetic energy can be presented in a quantized form as bursts of radiation with a quantized radiant energy, called photon Q , which is proportional to the frequency ν :

$$Q(\mathbf{r}, t) = h\nu \quad (6)$$

where h is the Planck constant ($h = 6.626 \cdot 10^{-34}$ J s). The quantum-matter interaction is mainly dependent on the frequency of the wave (i.e., its photon energy) and the energy level structure of the matter. As a wave interacts with a material in gas, liquid, or solid state, electrons, molecules, and/or nuclei are put into motion (i.e., rotation, vibration, or displacement) which leads to exchange of energy between the wave and the material.

The radiant energy, carried by the wave, is not delivered in a continuous deterministic way, but on a probabilistic quantum basis. The probability that a wave train will deliver its radiant energy at some place along the wave direction is proportional to the flux density of the wave at that place. If a very large number of wave trains are coexistent, then the overall average effect can be described by Maxwell's equations.

Radiation quantities

The *power flux density* of the electromagnetic radiation can be expressed by the Poynting vector $\mathbf{S}(\mathbf{r}, t)$ [W/m²], defined by

$$\begin{aligned}\mathbf{S}(\mathbf{r}, t) &= \mathbf{E}(\mathbf{r}, t) \times \mathbf{H}(\mathbf{r}, t) \\ &= \frac{|\mathbf{E}(\mathbf{r}, t)|^2}{\eta} \hat{\mathbf{r}} = S(\mathbf{r}, t) \hat{\mathbf{r}}\end{aligned}\quad (7)$$

where the last equality is true only if a TEM structure holds for the radiated field. The power flux density of a monochromatic field is directed along the wave propagation direction $\hat{\mathbf{r}}$.

In case of natural sources, the power flux density can be very low and angularly spread. It is convenient to express the radiation flow through the specific intensity or radiance (Chandrasekhar, 1960; Ishimaru, 1978). At a given point $\mathbf{r}$, the scalar specific intensity $I(\mathbf{r}, t)$ [W m⁻² sr⁻¹] (where sr stands for steradian or solid unit angle) represents the amount of elementary power dP [W] flowing within a solid angle $d\Omega$ through an elementary area dA [m²] oriented along the direction of the unit vector $\hat{\mathbf{n}}$ and is given by

$$I(\mathbf{r}, t) = \frac{dP(\mathbf{r}, t)}{dA \cos \theta d\Omega} = \frac{dS(\mathbf{r}, t)}{\cos \theta d\Omega} \quad (8)$$

where θ is the angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{r}}$. The specific intensity expresses the electromagnetic power leaving or intercepting an extended area in a given direction per unit project area in that direction. From Eq. 7, the *radiant intensity* is defined as $R(\mathbf{r}, t) = I(\mathbf{r}, t) dA \cos \theta$ [W sr⁻¹]. It is worth noting that both I and R are spatially invariant. Moreover, each radiation quantity can be considered as being composed by a number of sinusoidal waves or spectral components of frequency between ν and $\nu + d\nu$, each carrying a part of the power flux density. The spectral band over which these different components extend is called *spectral width* or *bandwidth* of the electromagnetic radiation. All radiation quantities have equivalent spectral quantities that correspond to their density per unit frequency (or wavelength) interval $d\nu$.

In random media, the specific intensity $I(\mathbf{r}, t)$ can be statistically related to the Poynting vector $\mathbf{S}$. If the latter at a given point $\mathbf{r}$ is a random function of time and each component of $\mathbf{S}$ varies in time, the tip vector of $\mathbf{S}$ moves randomly in time (Ishimaru, 1978). If this random variability is expressed by a probability density function, we can then define the specific intensity as the *ensemble*

average value of the random vector $\mathbf{S}$. This means that the $I(\mathbf{r}, t)$ is basically the sum of all random Poynting vectors $\mathbf{S}$ whose tips are located within a solid unit angle in the direction.

Electromagnetic radiation information

Electromagnetic radiation is employed in remote sensing as a means to sound and retrieve the features of an object and/or a medium (e.g., Elachi, 1988; Marzano and Visconti, 2002). In this respect, referring from simplicity to Eq. 4 or 5, peculiar features of the electromagnetic wave, perturbed in its interaction with matter, can be exploited:

1. *Amplitude*, described by the modulus of the wave field $|\mathbf{E}(\mathbf{r}, t)|$ or by its power flux density $S(\mathbf{r}, t)$ defined in Eq. 6. The transmitted wave amplitude can be, for example, modulated in order to perform a scatterer detection as in radars and lidars.
2. *Phase*, described by the phase function $\Psi(\mathbf{r}, \omega)$ in Eq. 4. The wave phase can be exploited, for example, by adopting frequency diversity (or multispectral techniques) for medium sounding as in radiometers and spectrometers, by measuring the time delay of the received signal to perform spatial ranging as in radars and lidars, by extracting the frequency shift of the received signal to estimate a scatterer radial velocity as in radars and lidars, by frequency modulating the transmitted signal to perform range profiling as in radars, and by coherently processing the received signal as in synthetic aperture radars.
3. *Polarization*, described by the polarization state vector $\mathbf{E}_0$ in Eq. 5. The wave polarization can be used, for example, by transmitting orthogonal states to detect scatterer shape and orientation as in radars and lidars and by receiving orthogonal states to detect emitter motion as in radiometers.

Some of the previous features of the electromagnetic radiation are analyzed in the following text.

Polarization diversity

An electromagnetic radiation consists of a coupled electric and magnetic force field. An electromagnetic wave that produces an electric field in a fixed plane along the direction of propagation is said to be linearly polarized (Ishimaru, 1991). In general, referring to plane waves for simplicity in Eq. 4, an *elliptically polarized* electrical field $\mathbf{E}(\mathbf{r}, t)$ can be represented, for example, by

$$\mathbf{E}(\mathbf{r}, t) = \text{Re}[e^{j\omega t} (E_{0h} \hat{\mathbf{h}} + E_{0v} \hat{\mathbf{v}}) e^{-j\beta \cdot \mathbf{r}}] \quad (9)$$

where $\hat{\mathbf{h}}$ and $\hat{\mathbf{v}}$ are the unit horizontal and vertical polarization vectors with E_{0h} and E_{0v} their complex amplitudes characterized by a modulus $|E_{0h}|$ and $|E_{0v}|$ and initial phase Ψ_{0h} and Ψ_{0v} , respectively. If $|E_{0h}| = |E_{0v}|$ and $\Delta\Psi_0 = \Psi_{0h} - \Psi_{0v} = \pm 90^\circ$, then the electromagnetic wave is *circularly polarized*; if $|E_{0h}| = 0$ or $|E_{0v}| = 0$ or $\Delta\Psi_0 = 0^\circ$, then the electromagnetic wave is

linearly polarized. The same definitions can be expressed through the related specific intensity by introducing the Stokes intensity vector $\mathbf{I}(\mathbf{r}, t)$ (Ishimaru, 1978).

If the polarization phase difference $\Delta\Psi_0$ is a random function, then the electromagnetic wave radiation does not have any clearly defined polarization and is called *unpolarized*; a typical example is the radiation from natural sources as the sun. In between, when the polarization phase difference $\Delta\Psi_0$ has a deterministic component (average) and random variability (fluctuation), the radiation is said to be *partially polarized*.

Radiation coherency

The wave phase function can be expressed through the *phase velocity*, the velocity at which a constant phase front progresses. From Eq. 4 or 5, putting $d\Psi(\mathbf{r}, \omega) = 0$, the phase velocity v_p is equal to

$$v_p = \frac{\omega}{\beta} \quad (10)$$

The wave phase velocity v_p is equal to the group velocity v_g , equal to $\partial\omega/\partial\beta$, only in case of a nondispersive medium such that ω is proportional to β .

If the medium is dispersive or has a narrow band $\Delta\nu$ around the central frequency ν_0 , then the instantaneous field will be a superposition of spectral components with a frequency between ν_0 and $\nu_0 \pm \Delta\nu/2$, whose random addition will lead to irregular fluctuations of the resultant field. The *coherency time* Δt is defined as the period over which there is a strong correlation of the field amplitude or the time after which two waves at ν and $\nu + \Delta\nu$ are out of phase by one cycle. From the phase function expression in a given point $\mathbf{r}$ and the equality $\omega\Delta t - (\omega + \Delta\omega)\Delta t = 1$, it results as follows:

$$\Delta t = \frac{2\pi}{\Delta\omega} = \frac{1}{\Delta\nu} = \frac{\Delta l}{v_p} \quad (11)$$

where $\Delta\nu$ is the *coherence bandwidth* and Δl is the *correlation length* at a given instant t .

The *wave interference* is a field amplitude effect. If two waves or two sources are coherent with each other, there is a systematic relationship between their instantaneous amplitudes. The amplitude of the resultant field varies between the sum (constructive interference) and the difference (destructive interference) of the two amplitudes. If the two waves are incoherent (i.e., infinite coherence bandwidth), then the correlation between the two waves is zero, and the power of the resultant wave is equal to the sum of the power of the two constituent waves. In general, an electromagnetic radiation will be *partially coherent*, and, if polarized, this will reflect in a partial polarization state.

Doppler effect

If the relative distance between a source radiating at a fixed frequency ν (or a wavelength λ) and an observer

varies, the electromagnetic wave, received by the observer, will have a frequency ν' (or a wavelength λ') different from ν . The difference $\nu_d = \nu - \nu'$ is called *Doppler frequency shift*. If the source–observer distance is decreasing, then $\nu_d > 0$; otherwise it holds $\nu_d < 0$.

The expression of the Doppler shift ν_d can be derived by differentiating the phase function $\Psi(\mathbf{r}, \omega)$ with respect to r so that from Eq. 5 we have as follows:

$$\begin{aligned} \nu_d &\equiv \frac{1}{2\pi} \frac{d\Psi(r, \omega)}{dr} = \pm u_r \frac{\beta}{2\pi} \\ &= \pm \frac{u \cos \theta_d}{\lambda} \\ &= \pm \nu \frac{\mathbf{u} \cdot \hat{\mathbf{r}}}{v_p} \end{aligned} \quad (12)$$

where $\mathbf{u}$ is the observer velocity vector and $\hat{\mathbf{r}}$ is the unit vector of the observer direction with θ_d the angle between $\mathbf{u}$ and $\hat{\mathbf{r}}$ and u_r the projection of $\mathbf{u}$ along $\hat{\mathbf{r}}$. In case of a *monostatic observation*, as in radars, where the observer and the source are collocated and a scattering object is moving, the expression of the Doppler shift ν_d in Eq. 11 must be multiplied by a factor 2 due to the two-way distance traveled by the electromagnetic radiation. It is clear that the Doppler effect can be used to retrieve target motion as in radars, but it is also useful to coherently process synthetic aperture radar signals to achieve very high-resolution imaging.

Radiation generation and detection

Electromagnetic radiation is generated by transformation of energy from other sources such as kinetic, chemical, thermal, electrical, magnetic, or nuclear. Several transformation mechanisms lead to electromagnetic waves over different regions of the electromagnetic spectrum. Generally speaking, the more organized (as opposed to random) the transformation mechanism is, the more coherent (i.e., narrower coherence bandwidth) is the generated radiation.

Electromagnetic fields, depending on the wavelength, can be generated in active remote sensing systems from periodic currents of electron charges in wires, electron beams, antenna surfaces, electron tubes, or molecular excitations followed by decay. The Rayleigh diffraction limit governs the beamwidth of the transmitted radiation such that it increases as the radiator size increases and the wavelength decreases. On the other hand, electromagnetic passive remote sensing systems measure the random emission of electromagnetic waves due to heat energy. The latter is the kinetic energy whose particle random motion results in excitation (electronic, vibrational, or rotational) due to random collisions followed by decay and emission. The transformation of heat energy into radiant energy is governed by the Planck's law.

In order to measure electromagnetic radiation, a collector is generally used followed by a detector. Radiation collectors may be either dipoles or reflectors at radio frequency and microwave frequency, whereas either

focusing lenses or reflectors in case of optical radiation. Detection is usually a transformation of electromagnetic radiation into another form of energy such as heat, electric current, or state change.

Bibliography

- Chandrasekhar, S., 1960. *Radiative Transfer*. New York: Dover.
- Elachi, C., 1987. *Introduction to the Physics and Technology of Remote Sensing*. New York: Wiley.
- Fung, A. K., 1994. *Microwave Scattering and Emission Models and their Applications*. Norwood, MA: Artech House.
- Ishimaru, A., 1978. *Wave Propagation and Scattering in Random Media*. New York: Academic, Vol. I and II.
- Ishimaru, A., 1991. *Electromagnetic Wave Propagation, Radiation and Scattering*. Englewood Cliffs, NJ: Academic Press/Prentice Hall.
- Marzano, F. S., and Visconti, G. (eds.), 2002. *Remote Sensing of Atmosphere and Ocean from Space: Models, Instruments and Techniques*. Dordrecht, NL: Kluwer. Advances in Global Change Research series, p. 246.
- Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing: Fundamentals and Radiometry*. Reading, MA: Addison Wesley, Vol. I.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)
[Fields and Radiation](#)
[Media, Electromagnetic Characteristics](#)
[Radiation, Multiple Scattering](#)
[Radiation, Polarization, and Coherence](#)
[Radiation, Volume Scattering](#)
[Radiative Transfer, Theory](#)
[Remote Sensing, Physics and Techniques](#)

RADIATION, GALACTIC, AND COSMIC BACKGROUND

David M. Le Vine
 Code 615, Cryospheric Sciences Branch, NASA/Goddard
 Space Flight Center, Greenbelt, MD, USA

Definition

Galactic radiation. Electromagnetic radiation whose sources lie primarily within our galaxy.
Cosmic background. Radiation associated with the origin of the universe (“big bang”).

Introduction

Stars and other celestial objects outside our solar system contribute radiation that can be important for Earth remote sensing. In many directions, this radiation is small. In fact, for many applications the “heavens” are used as a cold reference source (“cold sky”). However, this radiation can also be strong enough to require compensation as an unwanted background signal. For example, this is the case

in remote sensing of ocean salinity which is done at 1.4 GHz in the window set aside for radio astronomy applications. This background radiation can impact remote sensing of the Earth either as radiation directly to the sensor or as downwelling radiation that is reflected from the Earth surface to the sensor, and taking it into account can be complex because there are many sources of this celestial radiation both localized (discrete) and distributed in space.

Sources of radiation

There are three important categories of radiation that originate outside our solar system. These are the cosmic microwave background (CMB), line emission, and continuum emission such as is emitted by thermal sources.

Cosmic microwave background

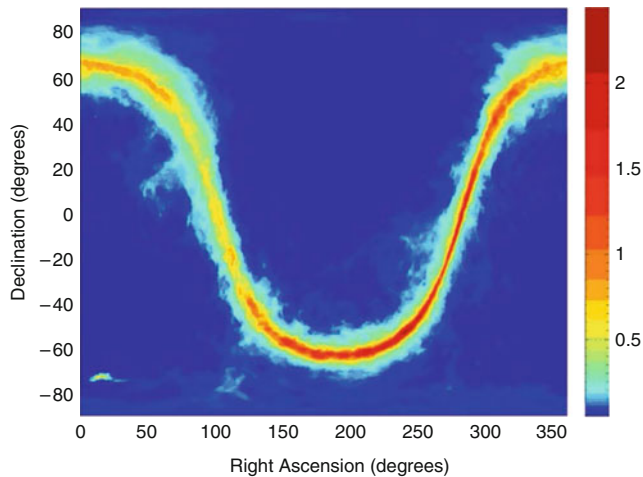
The cosmic microwave background radiation is a remnant of the origin of the universe in a “big bang.” It appears as a uniform (same in all directions) thermal source with an effective brightness temperature of 2.73 K. Although recent cosmological research has focused on the details of its spatial distribution, these variations (milli-Kelvin) are not important for most applications to remote sensing of the Earth. For microwave remote sensing applications, the cosmic background radiation is essentially constant in both space and time. This radiation can contribute to a measurement of an Earth-viewing sensor such as a radiometer observing the surface in a direct manner, for example, by entering through antenna side lobes in directions above the horizon. It also can contribute indirectly via reflection of downwelling radiation that is reflected off the surface (Blume and Kendall, 1982). The latter is especially important in remote sensing of targets such as the ocean surface where the reflection coefficient is relatively large and where applications such as monitoring sea surface salinity require high accuracy.

Line emission

Line emission is the radiation from matter associated with transitions in its atomic structure from one energy state to another with lower energy. The radiation tends to be relatively narrow in bandwidth (e.g., associated with a specific transition energy).

An example is the emission from a hyperfine transition in neutral hydrogen that occurs at 1.413 GHz. A window of the spectrum around this line (1.40–1.427 GHz) has been protected for passive measurements so that radio astronomers could use this radiation to monitor the distribution of hydrogen in the galaxy. This radiation provides information on the temperature, density, and motion of hydrogen. The radiation is concentrated around the plane of the galaxy, but clouds of hydrogen are widespread and no direction is observed without some such radiation.

Figure 1 shows the distribution of this radiation in celestial coordinates (an inertial coordinate system in which declination corresponds to latitude and right



Radiation, Galactic, and Cosmic Background, Figure 1 The equivalent brightness temperature of line emission from hydrogen at 1.413 GHz.

ascension corresponds to longitude referenced to a fixed time, defined by the “epoch”). The distribution of hydrogen lies largely in the plane of the galaxy. The distribution appears to have a U shape in celestial coordinates because the plane of the galaxy and the equatorial plane of the Earth, which is a reference plane in celestial coordinates, are tilted with respect to each other.

The data shown in [Figure 1](#) is a composite of two relatively recent radio astronomy surveys: (a) the Leiden/Dwingeloo survey by Hartmann and Burton (1997) which used the 25 m Dwingeloo radio telescope and covered the northern sky down to a declination of -30° , and (b) the IAR survey with the 30 m antenna at the Instituto Argentino de Radioastronomía (IAR) and reported by Arnal et al. (2000) which covered declinations south of -25° filling in the portions of the southern sky missing in the earlier survey. The result of the two surveys is data over the entire celestial sphere with spatial resolution of $0.5^\circ \times 0.5^\circ$ and good radiometric resolution ($\Delta T < 0.1$ K).

The data from these two surveys are plotted in [Figure 1](#) in the form of an equivalent brightness temperature. This is the temperature that a blackbody would have if it radiated the same power but distributed over a bandwidth of 20 MHz (Le Vine and Abraham, 2004). Any bandwidth could have been chosen, but 20 MHz is representative of the bandwidth of radiometers operating in the radio astronomy window at L-band.

Line emission provides a valuable tool in radio astronomy, and many spectral “windows” have been set aside for radio astronomy applications. For example, see the spectrum allocation chart at the National Telecommunications and Information Administration (NTIA, 2009).

Continuum radiation

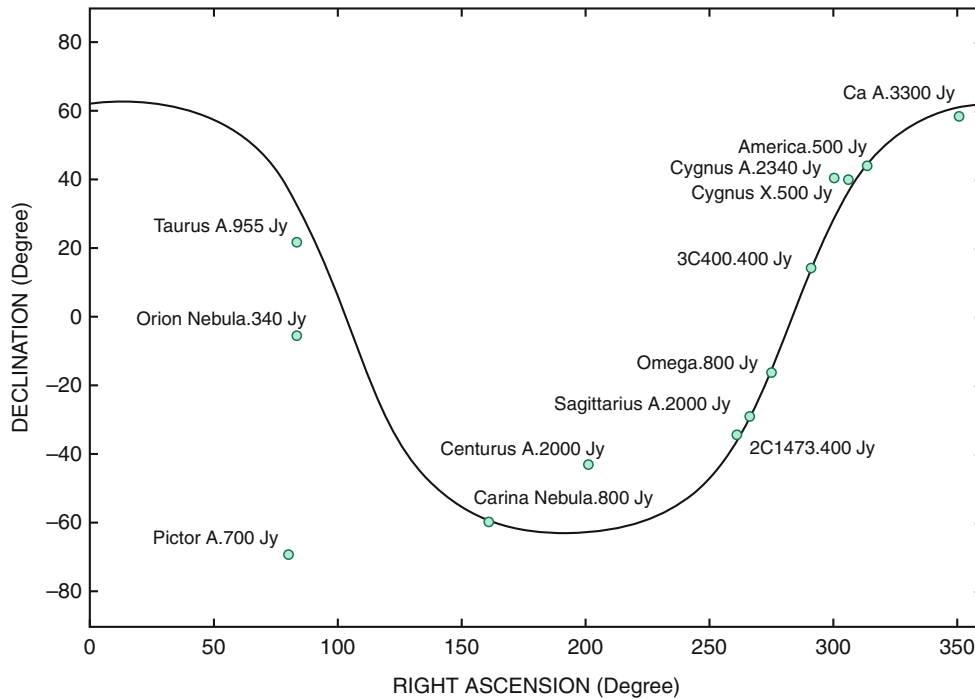
In addition to radiation with discrete spectra associated with atomic transitions, objects in the celestial sky emit

radiation with a continuous spectrum. The two dominant sources of this radiation are synchrotron radiation (radiation from accelerated relativistic electrons) and thermal radiation (radiation from a random collection of charged particles in motion). The latter has a spectrum given by Planck’s radiation law. The strongest sources of continuum radiation tend to lie in the plane of the galaxy and may be diffuse or discrete. Examples of discrete sources of synchrotron radiation, and among the strongest radio sources in the sky, are Cassiopeia A and the extragalactic source Cygnus A. The Orion Nebula is an example of a source of thermal radiation (Kraus, 1986). [Figure 2](#) shows the location of the strongest discrete sources together with their flux at 1.4 GHz in Jansky ($\text{W m}^{-2} \text{Hz}^{-1}$). (See Sect. 2.01.01.02 for a definition of flux.) The plot is in celestial coordinates and the solid line represents the plane of the galaxy. It is clear that most of the strong sources lie close to the plane of the galaxy.

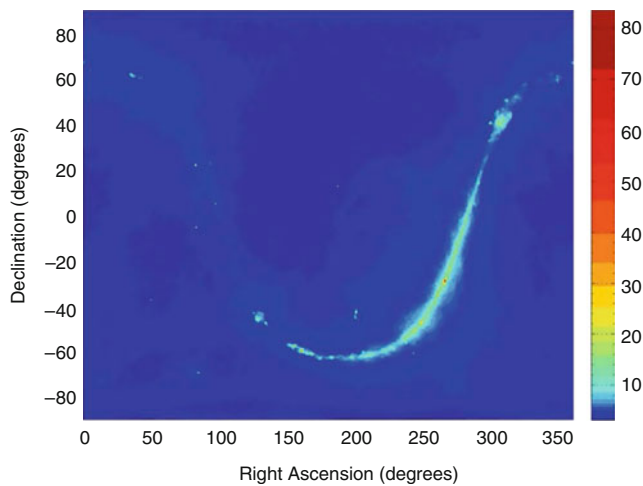
[Figure 3](#) shows a map of the continuum radiation at L-band. As in [Figure 1](#), the data is presented as the equivalent temperature of a blackbody with a bandwidth of 20 MHz centered at 1.413 GHz (Le Vine and Abraham, 2004). This map is a composite of two recent “all-sky” surveys of the continuum radiation at this frequency. The first was a survey with the Stockert telescope at Bonn University (Reich, 1982; Reich and Reich, 1986). The Stockert survey covered all of the northern sky and the southern sky to -19° declination. The data includes all sources except for a region around Cassiopeia A which was too strong for receivers used in this survey. The survey of the southern sky was completed using the 30 m radio telescope of the Instituto Argentino de Radioastronomía (Reich et al., 2001; Testori et al., 2001). The IAR continuum survey covers the southern sky for declination below -10° with spatial resolution and sensitivity similar to that of the Stockert survey. The composite map has a resolution of about $0.5^\circ \times 0.5^\circ$, an absolute accuracy of about 0.5 K, and a sensitivity of 0.05 K (rms error).

Notice the different scales on [Figures 1](#) and [3](#). The continuum radiation is significantly stronger than the line emission when spread over a 20 MHz bandwidth. Also, notice that again the radiation generally follows the plane of the galaxy.

[Figure 3](#) represents the radiation from all sources (e.g., distributed and discrete). Comparing with [Figure 2](#) many of the strongest discrete sources can be identified in [Figure 3](#). They would be easily identified if the data were plotted with higher resolution. [Figure 4](#) is an example. At the top is shown a section of [Figure 3](#) in the vicinity of Cygnus A plotted with higher resolution. Cygnus A is identified by the circle. The insert at the bottom of [Figure 4](#) is a plot of the same data but in three dimensions (vertical dimension is brightness temperature), to give an idea of the relative magnitude and size of Cygnus A. Cygnus A is a distant extragalactic object. The peak in radiation to the right of Cygnus A originates in the galaxy and corresponds to a line of sight along one of the spiral arms of the galaxy (Kraus, 1986).



Radiation, Galactic, and Cosmic Background, Figure 2 Celestial position of strong radio sources. The solid line indicates the galactic plane.



Radiation, Galactic, and Cosmic Background, Figure 3 The continuum radiation at L-band (1.413 GHz). This is a composite of the Stockert and IAR surveys.

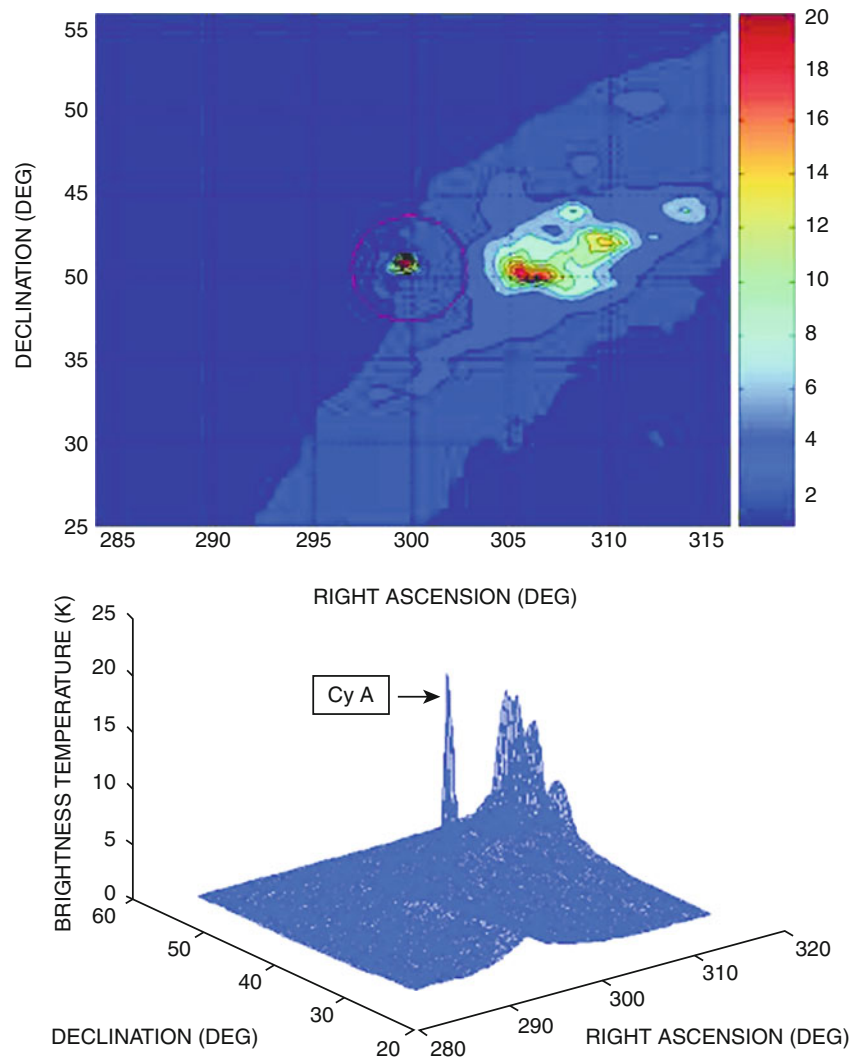
Radiation at other frequencies

The data presented in Figures 1 and 3 at L-band is available with a resolution of 0.5° at the Aquarius project website (Aquarius). The website includes explanations and references to the original surveys. Additional background can be found in Le Vine and Abraham (2004).

It would be convenient if the maps at L-band could be extrapolated to other frequencies. There are obvious problems. Clearly the line emission will be different for each frequency. The continuum data (Figure 3) could be extrapolated if the spectral shape were known. For example, if the source were thermal, with a fixed temperature, the radiated power would be described by the Rayleigh-Jeans approximation (i.e., vary as f^2). Unfortunately, there are many processes responsible for the observed radiation, and for most sources the radiated power decreases with frequency (Kraus, 1986). Reich and Reich (1988) used the survey described above (1.4 GHz) and an all-sky survey at 480 MHz (Haslam et al., 1982) to compute an index, β , for brightness temperature ($T_B \sim f^{-\beta}$). They obtained values of $2.3 < \beta < 3.0$ with most values around $\beta = 2.7$. Whether or not this can be used to reliably scale the L-band data needs to be verified by comparison with measurements.

Summary

Radiation from outside of our solar system can impact an Earth-viewing sensor either as direct radiation or radiation reflected from the Earth surface to the sensor. In the microwave part of the spectrum, there are three important categories of this radiation: the cosmic microwave background (a uniform thermal source at 2.73 K), line emission, and continuum emission. The hydrogen line at 1.413 GHz (L-band) is an example of line emission, and



Radiation, Galactic, and Cosmic Background, Figure 4 Top: expanded section of Figure 2 in the vicinity of Cygnus A. Bottom: the data at the top plotted in three dimensions.

synchrotron emission is an example of continuum emission. Strong radio sources tend to be concentrated near the plane of the galaxy.

Bibliography

- Aquarius: Available from World Wide Web: <http://aquarius.nasa.gov/>
- Arnal, E. M., Bajaja, E., Larrarte, J. J., Morras, R., and Pöppel, W. G. L., 2000. A high sensitivity HI survey of the sky at $\delta \leq -25^\circ$. *Astronomy and Astrophysics Supplement Series*, **142**, 35–40.
- Blume, H.-J. C., and Kendall, B. M., 1982. Passive microwave measurements of temperature and salinity in coastal zones. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-20**(3), 394–404.
- Hartmann, D., and Burton, W. B., 1997. *Atlas of Galactic Neutral Hydrogen*. Cambridge, UK: Cambridge University Press.
- Haslam, C. G. T., Salter, C. J., Stoffel, H., and Wilson, W. E., 1982. A 408 MHz all-sky continuum survey II. The atlas of contour maps. *Astronomy and Astrophysics Supplement Series*, **4**, 1–143.
- Kraus, J. D., 1986. *Radio Astronomy*. New York: McGraw-Hill.
- Le Vine, D. M., and Abraham, S., 2004. Galactic noise and passive microwave remote sensing from space at L-band. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(1), 119–129.
- NTIA, 2009. National Telecommunications and Information Administration: Available from World Wide Web: <http://www.ntia.doc.gov/osmhome/allochrt.html>.
- Reich, W., 1982. A radio continuum survey of the Northern Sky at 1420 MHz – Part 1. *Astronomy and Astrophysics Supplement Series*, **48**, 219–297.
- Reich, P., and Reich, W., 1986. A radio continuum survey of the Northern Sky at 1420 MHz – Part II. *Astronomy and Astrophysics Supplement Series*, **63**, 205–292.
- Reich, P., and Reich, W., 1988. A map of spectral indices of the galactic radio continuum emissions between 408 MHz and

- 1420 MHz for the entire Northern Sky. *Astronomy and Astrophysics, Supplement Series*, **74**, 7–23.
- Reich, P., Testori, J. C., and Reich, W., 2001. A radio continuum survey of the Southern Sky at 1420 MHz. The atlas of contour maps. *Astronomy and Astrophysics*, **376**, 861–877.
- Testori, J. C., Reich, P., Bava, J. A., Colomb, F. R., Hurrel, E. E., Larrarte, J. J., Reich, W., and Sanz, A. J., 2001. A radio continuum survey of the Southern Sky at 1420 MHz: observations and data reduction. *Astronomy and Astrophysics*, **368**, 1123–1132.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)
[Radiation, Electromagnetic](#)

RADIATION, MULTIPLE SCATTERING

Frank S. Marzano
 Department of Information Engineering, Sapienza
 University of Rome, Rome, Italy
 Centre of Excellence CETEMPS, University of L'Aquila,
 L'Aquila, Italy

Definition

Multiple scattering is an effect due to propagation of waves through a medium characterized by a random distribution of discrete scatterers, usually embedded in a continuum medium. The random distribution may be due to the size probability distribution of the scatterer polydispersion, their orientation, and their relative motion, whereas scatterers may be particles, objects, or medium spatial irregularities.

Introduction

The propagation of waves through a medium characterized by a random distribution of discrete scatterers, usually embedded in a continuum medium, is usually characterized by multiple scattering. This effect is strongly dependent on the ratio between the scatterer average size and the wavelength. Examples of such media, where multiple scattering may happen, are hydrometeors and aerosols in atmosphere; leaves, branches, and trunks in forests; ice-air irregularities in snow packs; air bubbles and fishes in oceans; and cellular structures and blood cells in biology. The multiple scattering effect can be macroscopically referred to either electromagnetic waves or acoustic waves.

Multiple scattering and radiative transfer

In order to identify multiple scattering effects, it is customary to distinguish between two extreme cases: *sparse* and *dense random distribution* of particles. This distinction can be first evaluated by quantifying the fractional volume occupied by the ensemble of particles: in case it is less than few tenths percent, then we deal with *sparse distribution* of particles, whereas a fraction volume larger than few percent represents *dense distribution* of particles.

Intermediate cases may be generally labeled as *rarefied distribution* of particles.

Indeed, more rigorously, the distinction of the random distribution regime should be made on the basis of the optical distance and the albedo of particle distribution which express, respectively, the properties of medium extinction and geometrical extension and scattering efficiency with respect to total extinction (equal to the sum of absorption and scattering). Both medium optical distance and albedo depend on the wavelength, scatterer size, composition, and orientation, as well as on the transmitter and receiver characteristics. When the medium optical thickness is low, the medium is said to be *tenuous*, otherwise *non-tenuous*.

In this respect, it is convenient to adopt the *radiative transfer theory* as a general framework to describe both absorption and scattering of wave propagation through a random medium with discrete scatterers and then distinguish among its various approximations. A fairly general expression of the radiative transfer equation (RTE), for the unpolarized diffuse specific intensity $I_d(\mathbf{r}, \hat{\mathbf{s}})$ ($\text{W m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}$) and under the basic assumption of *independent wave scattering* from each scatterer, is given by Tsang et al. (1985) and Ishimaru (1997):

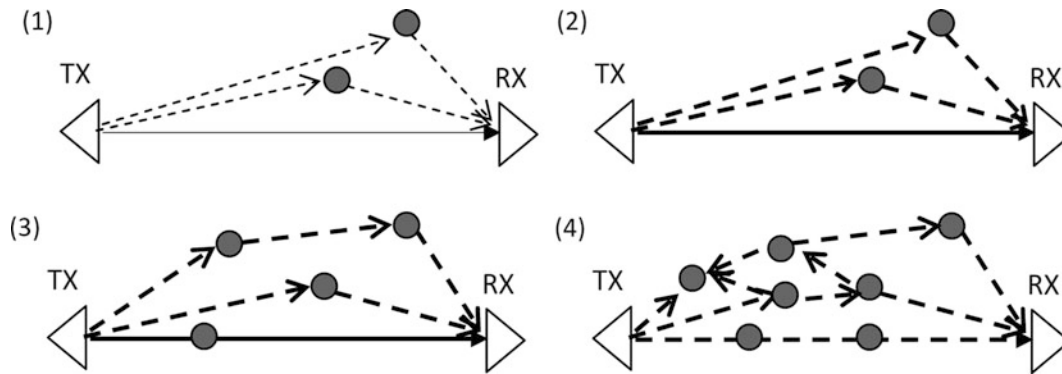
$$\frac{dI_d(\mathbf{r}, \hat{\mathbf{s}})}{d\tau} = -I_d(\mathbf{r}, \hat{\mathbf{s}}) + \frac{w(\mathbf{r})}{4\pi} \times \int_{4\pi} p(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}') I_d(\mathbf{r}, \hat{\mathbf{s}}) d\Omega' + J(\mathbf{r}, \hat{\mathbf{s}}) \quad (1)$$

where τ is the medium optical thickness or distance (product of the distance ds and extinction coefficient k_e), $\mathbf{r}$ is the position vector and $\hat{\mathbf{s}}$ is the direction unit vector, w is the albedo (ratio between medium scattering coefficient k_s and extinction coefficient k_e), p is the scattering phase function, Ω is the solid angle, and J is the source function due to *source intensity* in the form of either medium thermal emission or attenuated incident wave (or reduced specific intensity). The total intensity $I(\mathbf{r}, \hat{\mathbf{s}})$ is given by the sum of the diffuse specific intensity $I_d(\mathbf{r}, \hat{\mathbf{s}})$ and the coherent one $I_c(\mathbf{r}, \hat{\mathbf{s}})$.

The RTE describes the spatial differential variation of the specific intensity which is decreased by extinction (first term on the right side), increased by the multiple scattering mechanism (first term on the right side, often called *pseudo-source function*), and increased by the emitted and incident radiation (third term on the right side). Thus, the analysis of the RTE indicates that multiple scattering is an effect which tends to recover radiation due to double, triple, etc., order of scattering of the radiation itself (Marzano and Roberti, 2003).

Wave scattering from random distribution of particles

The radiative transfer framework can provide a useful way to distinguish among the various random distribution



Radiation, Multiple Scattering, Figure 1 Schematic view of multiple scattering regimes: (1) single scattering, (2) first-order multiple scattering, (3) multiple scattering, and (4) diffusive scattering. TX stands for radiation transmitter (or source) and RX for receiver. Particles are drawn as circles, and dashed lines stand for scattered waves (and when bold, it means that they are also attenuated), whereas continuous lines denote direct wave (and when bold, it means that they are also attenuated).

scenarios. A schematic view of the basic multiple scattering mechanisms is shown in Figure 1.

In case of a *sparse tenuous random distribution*, we can use the so-called single scattering approximation where the incident wave, irradiated by the transmitter, reaches the receiver after encountering very few particles. In this case, the scattered wave is assumed to be due to a single scattering by a particle and all higher-order scattering are considered to be negligible. This means that the optical thickness is much than one, and the specific intensity in the multiple scattering term of Equation 1 is approximated by the source-specific intensity. The “single scattering” approximation can be adopted in many environmental applications such as weather radars in case of electromagnetic waves and ocean acoustics in case of acoustic waves (e.g., Tsang et al., 1985; Stephens, 1970).

If the particle density is increased but can be still considered a sparse distribution, it is needed to take into account both absorption and scattering along the path. This implies to deal with a *sparse non-tenuous random distribution* of scatterers. In this case, the scattered wave is scattered once by a particle, but both the incident and scattered waves have likewise been attenuated. This means that the optical thickness is not negligible and the specific intensity in the multiple scattering term of Equation 1 is still approximated by the source-specific intensity. This approximation is very often called *first-order multiple scattering* in order to point out that some of the multiple scattering is taken into consideration. Microwave extinction due to rain and optical extinction due to aerosols, for example, have been extensively studied using this approximation (e.g., Klett, 1981; Marzano and Ferrauto, 2003).

When the particle density is such that a *dense tenuous distribution* is considered, we cannot disregard any more severe multiple scattering effects even though independent scattering from each particle is still assumed. This problem can be described by the so-called diffusive

scattering (or diffusion approximation), where the scattered wave is almost uniformly diffracted in all directions. When the scattering is angularly uniform, diffusive scattering is called *isotropic scattering*. The latter corresponds to a constant scattering phase function p in Equation 1. This approximation has been successfully applied, for example, in the analysis of optical fiber oximeter of blood (e.g., Johnson and Guy, 1972).

When neither *first-order multiple scattering* nor *diffusive scattering* can be applied, we need to deal with a general random distribution of particles whose fractional volume is around 1 % typically. Such distributions may be classified as *rarefied non-tenuous media*. This means that the independent wave scattering is still valid and the problem needs to be solved resorting to the complete integrodifferential expression of the RTE, given in Equation 1. In this case, not only all order of multiple scattering and absorption effects must be considered but also the scattered wave is angularly not uniform. These applications range, for example, from microwave radiometry of clouds and precipitation to radar meteorology and microwave scattering from vegetated canopies (e.g., Ishimaru and Cheung, 1980; Ferrazzoli and Guerriero, 1996; Marzano et al., 2003; Marzano, 2007).

Problems dealing with *dense non-tenuous distribution media*, where the spatial correlation of particle wave scattering is not negligible, cannot be approached by means of the RTE model (Tsang and Ishimaru, 1987). The lack of the independent scattering assumption prompts to face the *analytical theory of multiple scattering*.

Analytical theory of multiple scattering

As already mentioned, the radiative transfer theory describes the power intensity balance of the transmission, absorption, and scattering processes and is heuristically derived from considerations on energy conservation, as apparent in Equation 2.

The same problem can be cast in terms of field quantities and their statistical characterization. When this approach is followed, we can refer to the *analytical theory of multiple scattering* (or multiple scattering theory), whose general formulation is well summarized by the integral equation set of Twersky-Foldy (Twersky, 1964; Ishimaru, 1997).

The *analytical theory of multiple scattering* generally starts from the Maxwell wave equation of a scalar field $\Psi(\mathbf{r})$, which can represent either a component of the electric or magnetic field or a pressure wave and describes the various processes of multiple scattering in a detailed way, usually resorting to the concise formalism of the diagram method. The general result of it is that the total field $\Psi(\mathbf{r})$ is composed by the following: (1) the incident wave, (2) multiply scattered waves involving chains of successive scattering through different scatterers, and (3) multiply scattered waves containing all the paths which go through a scatterer more than once. This general picture is consistent with that schematically shown in Figure 1. The *Twersky theory* adopts the so-called expanded representation which retains only the previous groups (1) and (2) but disregards the effects due to group (3). When the number of scattering events becomes large, the errors associated with the Twersky approximation become very small.

Starting from the Twersky expanded representation, it is then possible to provide a statistical description of the random field $\Psi(\mathbf{r})$, distinguishing between the coherent (or average) field $\Psi_c(\mathbf{r})$ and the diffuse (also incoherent or fluctuating) field $\Psi_d(\mathbf{r})$:

$$\Psi(\mathbf{r}) = \Psi_c(\mathbf{r}) + \Psi_d(\mathbf{r}) \quad (2)$$

where $\Psi_c(\mathbf{r}) = \langle \Psi(\mathbf{r}) \rangle$ and $\langle \Psi_d(\mathbf{r}) \rangle = 0$ with the angle brackets indicating ensemble average. It can be shown that the total field intensity $\langle |\Psi(\mathbf{r})|^2 \rangle$, which can be decomposed into the sum of coherent intensity $|\langle \Psi(\mathbf{r}) \rangle|^2$ and diffusive intensity $\langle |\Psi_d(\mathbf{r})|^2 \rangle$ from Equation 2, satisfies the so-called Twersky-Foldy integral equations. The latter are, in their turn, consistent with the first-order smoothing approximation of the more rigorous Dyson and Bethe-Salpeter equations (e.g., Furutsu, 1975).

Considering that the RTE in Equation 1 and the Twersky-Foldy integral equations describe the same multiple scattering phenomenon, even though from two different points of view, a link between the two theories is expected (Ishimaru, 1975). Indeed, if $\Gamma_\Psi(\mathbf{r}, \Delta\mathbf{r})$ is the correlation function (or mutual coherence function) of the random field $\Psi(\mathbf{r})$ in a given position $\mathbf{r}$ and at a distance $\Delta\mathbf{r}$, it can be shown that it holds the following approximate equality (Furutsu, 1975):

$$\Gamma_\Psi(\mathbf{r}, 0) = \langle |\Psi(\mathbf{r})|^2 \rangle \cong \int_{4\pi} I(\mathbf{r}, \hat{\mathbf{s}}) d\Omega \quad (3)$$

The previous relation indicates that the angular integral of the total specific intensity I (i.e., the radiation intensity)

is intensity $\langle |\Psi(\mathbf{r})|^2 \rangle$. Equation 3 can be generalized showing that the field correlation function Γ_Ψ is related to the Fourier transform of the specific intensity I . It shows that the description of multiple scattering effects, based on the radiative transfer theory, does include information concerning the field quantities in the form of mutual coherence function (Ishimaru, 1997).

In a dense non-tenuous medium, the assumption of independent scattering, which is often used in conventional radiative transfer theory, is not valid (Tsang and Ishimaru, 1987). In the ladder approximation of correlated scatterers, wave interactions between different particles (whether of far, near, or intermediate range) are weighted by the pair distribution function of particle positions; therefore, the wave interactions of all ranges of distance separation are included. The extinction rate, albedo, and phase matrix of the dense medium radiative transfer equation are expressed in terms of the physical parameters of the medium. The dense medium radiative transfer equation takes the following into account: (a) the scattering by correlated scatterers, (b) the pair distribution function of scatterer positions, and (c) the effective propagation constant of a dense medium. In particular, for volume fractions higher than 10 %, the quasicrystalline approximation has shown to be effective solutions when coupled either with the Percus-Yevick pair correlation function or with the coherent potential approximation (Tsang et al., 1985).

Bibliography

- Ferrazzoli, P., and Guerriero, L., 1996. Passive microwave remote sensing of forests: a model investigation. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 433–443.
- Furutsu, K., 1975. Multiple scattering of waves in a medium of randomly distributed particles and derivation of the transport equation. *Radio Science*, **10**, 29–44.
- Ishimaru, A., 1975. Correlation function of a wave in a random distribution of stationary and moving scatterers. *Radio Science*, **10**, 45–52.
- Ishimaru, A., 1997. *Wave Propagation and Scattering in Random Media*. Piscataway/Oxford: IEEE Press/Oxford University Press.
- Ishimaru, A., and Cheung, R. L.-T., 1980. Multiple-scattering effect on radiometric determination of rain attenuation at millimeter-wavelengths. *Radio Science*, **15**, 507–516.
- Johnson, C. C., and Guy, A. W., 1972. Nonionizing electromagnetic wave effects in biological material and systems. *Proceedings of the IEEE*, **60**, 692–718.
- Klett, J. D., 1981. Stable analytical inversion solution for processing lidar returns. *Applied Optics*, **20**, 211–220.
- Marzano, F. S., 2007. Predicting antenna noise temperature due to rain clouds at microwave and millimeter-wave frequencies. *IEEE Transactions on Antennas and Propagation*, **55**, 2022–2031.
- Marzano, F. S., and Ferrauto, G., 2003. Relation between the radar equation and the first-order backscattering theory. *Atmospheric Chemistry and Physics*, **3**, 813–821.
- Marzano, F. S., and Roberti, L., 2003. Numerical investigation of intense rainfall effects on coherent and incoherent slant-path propagation at K band and above. *IEEE Transactions on Antennas and Propagation*, **41**, 965–977.
- Marzano, F. S., Roberti, L., Di Michele, S., Tassa, A., and Mugnai, A., 2003. Modeling of apparent radar reflectivity due to

- convective clouds at attenuating wavelengths. *Radio Science*, **38**, 1002, doi:10.1029/2002RS002613.
- Stephens, R. W. B., 1970. *Underwater Acoustics*. New York: Wiley Interscience.
- Tsang, L., 1987. Passive remote sensing of dense non-sparse media. *Journal of Electromagnetic Waves and Applications*, **1**, 159–173.
- Tsang, L., and Ishimaru, A., 1987. Radiative wave equations for vector electromagnetic propagation in dense nonsparse media. *Journal of Electromagnetic Waves and Applications*, **1**, 52–72.
- Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley.
- Twersky, V., 1964. On propagation in random media of discrete scatterers. Stochastic processes in mathematical physics and engineering. *American Mathematical Society*, **84**, 84–116.
- Wen, B., Tsang, L., Winebrenner, D. P., and Ishimaru, A., 1990. A Dense medium radiative transfer theory: comparison with experiment and application to microwave remote sensing and polarimetry. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 46–59.

Cross-references

Electromagnetic Theory and Wave Propagation
Fields and Radiation
Media, Electromagnetic Characteristics
Optical/Infrared, Radiative Transfer
Radiation, Volume Scattering
Radiative Transfer, Theory

RADIATION, POLARIZATION, AND COHERENCE

Yang Du
Zhejiang University, Hangzhou, People's Republic of China

Definition

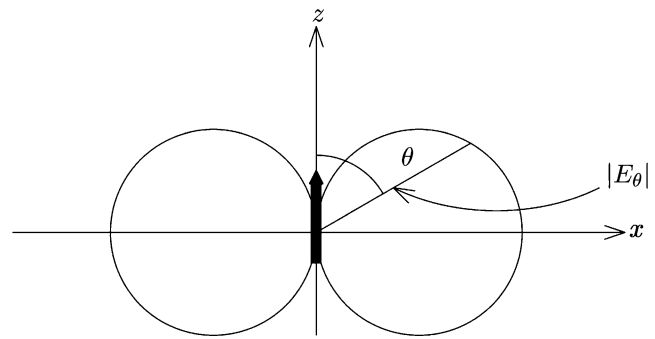
Radiation. The process of electromagnetic energy propagation from the source.

Polarization. A property of electromagnetic waves to describe the orientation of their oscillations.

Coherence. A property of electromagnetic waves that enables stationary interference.

Radiation

Radiation of energy characterizes a source in an unbounded space. To illustrate the radiation mechanisms and the dependence of the radiation on the source, first consider the case of a single wire, where there is no radiation if a charge is not moving or charge is moving with uniform velocity along a straight and infinite wire, whereas there is radiation if the wire is deformed in certain ways (curved, bent, truncated, discontinuous) or charge is oscillating (Balanis, 1997). Next consider another simple case of current I forming an electric dipole of moment Il by extending over an incremental length l , the results of which can be readily extended to general source distributions. For simplicity, current element is placed at origin and directed in the z direction. The magnetic



Radiation, Polarization, and Coherence, Figure 1 The magnitude of the radiated field $|E_\theta|$ as a function of θ .

vector potential, having its rectangular components corresponding to those of sources, is z directed. The vector Helmholtz equation governing the magnetic vector potential is reduced to a scalar one governing the z component of the magnetic vector potential. The solution is simply $A_z = \frac{Il}{4\pi r} e^{ikr}$. The electric and magnetic fields can be readily obtained from the magnetic vector potential (Harrington, 1961). The electromagnetic fields behave very differently as the observation point moves away from the source. For the case that the source dimension l is very small compared to the wavelength λ , three spatial regions can be classified:

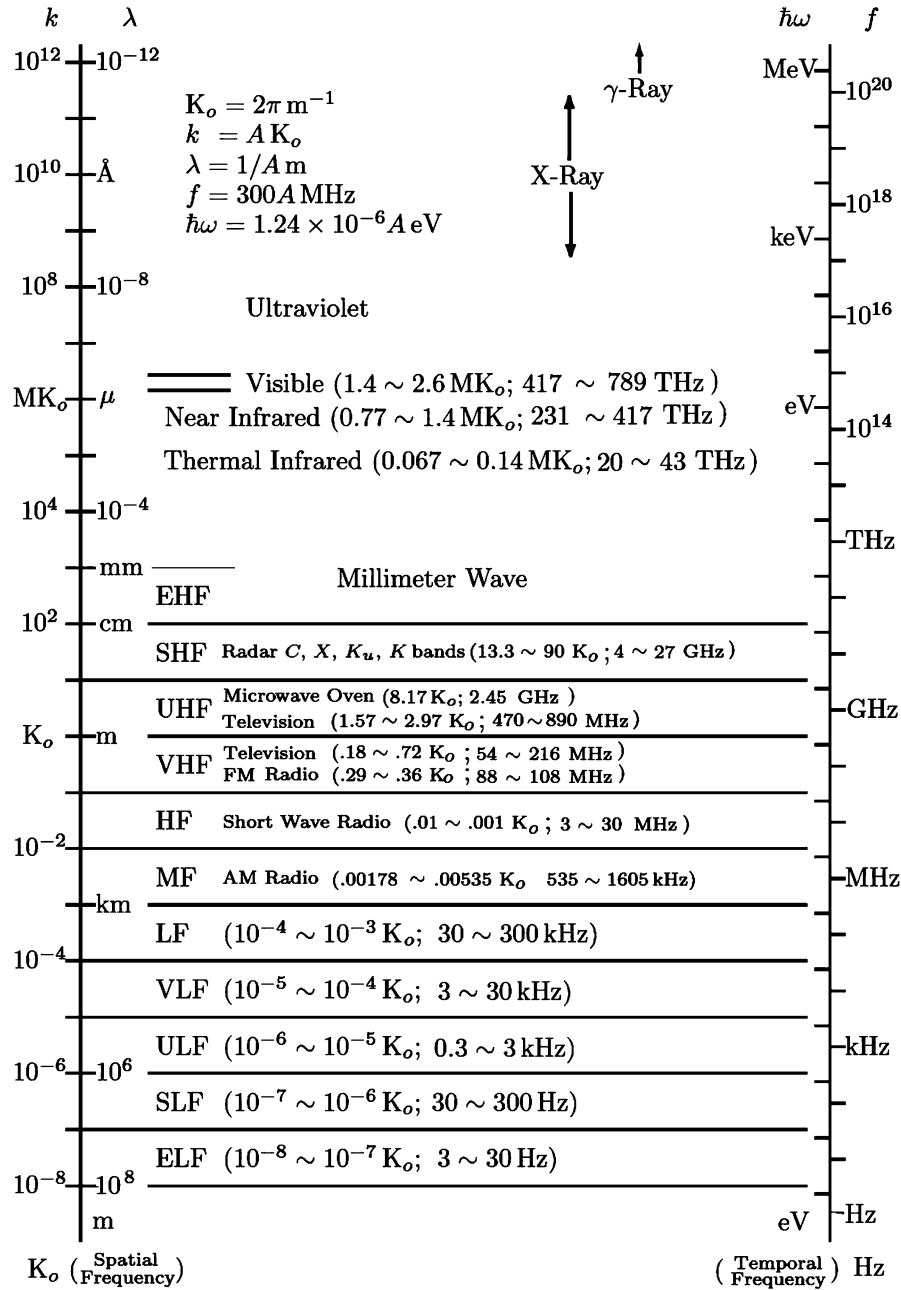
- The near zone (static), if $l \ll r \ll \lambda$
- The intermediate zone (induction), if $l \ll r \sim \lambda$
- The far zone (radiation), if $r \gg \lambda$

In the radiation zone, the fields fall off as r^{-1} and are transverse to the radial direction (Jackson, 1998). The radiation field is depicted in Figure 1, where the magnitude of $|E_\theta|$ as a function of θ is shown.

For general current source $\mathbf{J}(\mathbf{r}, \omega)$, the magnetic vector potential is obtained through the Green's function. In characterizing radiated fields in the far zone, if the source dimensions are small compared to the wavelength, Taylor series expansion of the exponential factor of the integrand can be performed, with each term in the series attaining certain physical interpretation, such as electric dipole, magnetic dipole, and electric quadrupole, for the first few terms. A more systematic treatment can be carried out by means of multipole expansion (Jackson, 1998). According to the wavelength, a diagram of the electromagnetic spectrum is shown in Figure 2.

Polarization

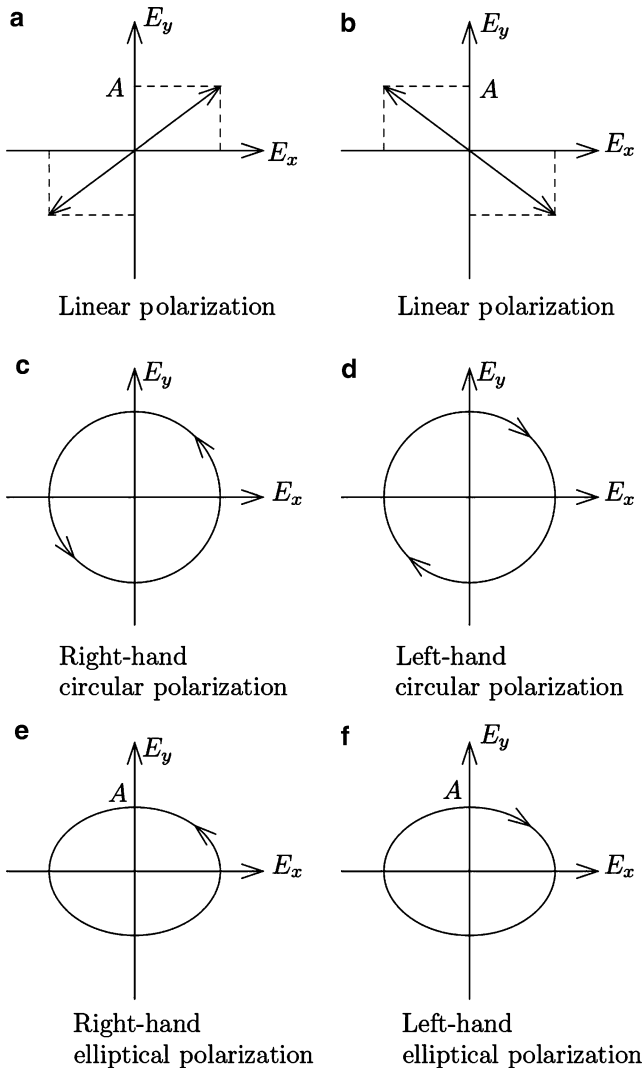
Polarization is a property of electromagnetic waves to describe the orientation of their oscillations. Polarization may be helpful in understanding the physics of some elusive phenomena. For example, the sources of the gamma-ray bursts (GRBs) remain a mystery albeit they are known to locate in distant galaxies. The detection of polarization may provide clues on the sources' identities and on the gamma-ray-generating mechanism (Waxman, 2003).



Radiation, Polarization, and Coherence, Figure 2 Electromagnetic wave spectrum (Source: J.A. Kong, *Electromagnetic Wave Theory*, EMW Publishing, Cambridge, MA, 2005. Reprinted with permission of J.A. Kong).

For a plane wave, the polarization at a fixed point of space is associated with the time variation of the tip of the electric field vector in a plane perpendicular to the direction of propagation. If the two mutually orthogonal components of the electric vector in the plane have the same or opposite phases, then the tip moves along a straight line and the wave is linearly polarized. If the two components have different phases, the locus of the tip is, in general, an ellipse and the wave is elliptically

polarized, except the special case when the magnitudes of the two components are the same and the phase difference is $\pm 90^\circ$, the locus of the tip is a circle, and the wave is circularly polarized. Furthermore, for an elliptically or circularly polarized wave, if the right-hand thumb points to the propagation direction and the fingers are along the tip sweeping direction, the wave is right-hand polarized; otherwise, it is left-hand polarized (Jackson, 1998; Kong, 2005). These polarizations are sketched in Figure 3.



Radiation, Polarization, and Coherence, Figure 3 Sketch of polarizations (Source: J.A. Kong, *Electromagnetic Wave Theory*, EMW Publishing, Cambridge, MA, 2005. Reprinted with permission of J.A. Kong).

To determine the state of polarization of a plane wave from observations, the four Stokes parameters are useful. These parameters, being quadratic in fields, allow for mere intensity measurements. If the electric field is written as $\mathbf{E}(t) = \hat{h}E_h + \hat{v}E_v$, where $\hat{h}$ and $\hat{v}$ are mutually orthogonal unit vectors perpendicular to the direction of propagation, the four Stokes parameters are defined as

$$\begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \frac{1}{\eta} \begin{bmatrix} |E_v|^2 + |E_h|^2 \\ |E_v|^2 - |E_h|^2 \\ 2 \operatorname{Re}(E_v E_h^*) \\ 2 \operatorname{Im}(E_v E_h^*) \end{bmatrix},$$

where η is the wave impedance of the medium, superscript asterisk refers to conjugate operation, and

$\operatorname{Re}()$ and $\operatorname{Im}()$ refer to real and imaginary parts (Tsang et al., 1985; Kong, 2005).

In the above the concept of full polarization is applied to monochromatic field. Actually, monochromaticity is not a necessary condition for the field to be fully polarized. The requirement is that the ratio of two mutually orthogonal components of the electric field perpendicular to the propagation direction is constant (Mandel and Wolf, 1995).

In the case of random electromagnetic fields, the concept of polarization is closely related to the concept of correlation between components of the electromagnetic fields, just as is the concept of coherence. For this reason, we defer the description of polarization to the coherence subsection so as to bring forth an integrated treatment.

Coherence

Coherence is a property of electromagnetic waves that enables stationary interference (e.g., as those patterns observed in the famous Young's slit interference experiment). More generally, coherence describes all correlation properties between physical quantities of a wave.

Temporal coherence between two electromagnetic fields refers to the capability of producing interference fringes in an interferometer (e.g., Michelson interferometer for lights) even if a time delay is introduced between them. A measure of the temporal coherence is the coherent time, which is approximately the reciprocity of the effective bandwidth of the field (Mandel and Wolf, 1965).

The electromagnetic fields may be generated by random sources and become random fields. At times, the resultant randomness is beneficial to the specific application. For instance, light beams generated by sources that are partially coherent are useful in the reduction of speckle in optical systems. If the fields are generated by a Gaussian random process such as thermal sources, for example, incandescent matter, then the second-order tensors, such as the concept of correlation between the fields at two space-time points and the governing dynamic laws, are adequate for the description of fields. The concepts of coherence and polarization can attain their respective meaning in terms of the concept of correlation.

The 2×2 electric cross-spectral tensor of the field is defined as the correlation between the components perpendicular to the propagation direction of electric fields at two positions. The coherence matrix at some point is determined by making the two positions identical in the electric cross-spectral tensor. The degree of polarization and degree of coherence can be subsequently defined. The former is defined as the square root of one minus four times the determinant of the coherence matrix divided by squared value of the trace of the coherence matrix, while the latter is defined as the ratio of the trace of the electric cross-spectral tensor over the square root of the product of the traces of the electric cross-spectral tensor specified at each point (James, 1994; Wolf, 2003; Ellis et al., 2004).

The dynamics of polarization and coherence of random electromagnetic fields can be quite different from their

deterministic counterparts. For the simple case of a partially coherent, partially polarized light beam that propagates along the same axis, analytical results show that the degree of polarization of the light beam changes as the beam propagates in free space (James, 1994). This observation is contrary to the assumption about the invariance of the state of polarization of a light beam propagating in free space. In fact, the quantities that change are not limited to the state of polarization; the state of coherence changes as well. Moreover, these coherence and polarization properties can be characterized for an electromagnetic beam propagating in free space as well as in linear media. In the latter case, the media can be deterministic or random (Wolf, 2003).

However, the issue of inclusion of the vectorial character of electromagnetic waves in defining coherence is complex and delicate. There are alternative definitions of the degree of coherence (Tervo et al., 2003; Setälä et al., 2004). Searching for the right form of the degree of coherence is an ongoing endeavor (Luis, 2007).

Meanwhile, there are voices that advocate a more comprehensive treatment by arguing that magnetic field and magnetic coherence are important for tightly focused electromagnetic waves and magnetic interactions in semiconductor quantum dots. Extension of electric coherence to magnetic coherence, and connection between these two coherences in free electromagnetic fields, is proposed (Jouttenus et al., 2005).

There are also other issues to deal with. The above descriptions are based on the assumption that the random electromagnetic fields are generated by Gaussian random process. However, there are fields that are generated by non-Gaussian random process, as evident in some important experiments of Hanbury Brown and Twiss, and in optics with the development of some new types of light source, for example, the Smith-Purcell radiator and the optical maser. In these cases, correlation phenomena need to be readdressed (Mandel and Wolf, 1965).

Summary

Radiation of energy characterizes a source in an unbounded space. In the radiation zone, the fields fall off as the reciprocity of distance from the source and are transverse to the radial direction. Polarization is a property of electromagnetic waves to describe the orientation of their oscillations. Coherence is a property of electromagnetic waves that enables stationary interference. In the case of random electromagnetic fields, the concepts of polarization and coherence are closely related to the concept of correlation between components of the electromagnetic fields.

Bibliography

- Balanis, C. A., 1997. *Antenna Theory: Analysis and Design*, 2nd edn. Hoboken, NJ: Wiley.
- Ellis, J., Dogariu, A., Ponomarenko, S., and Wolf, E., 2004. Correlation matrix of a completely polarized, statistically stationary electromagnetic field. *Optics Letters*, **29**, 1536.

- Harrington, R. F., 1961. *Time-Harmonic Electromagnetic Fields*. New York: McGraw-Hill.
- Jackson, J. D., 1998. *Classical Electrodynamics*, 3rd edn. New York: Wiley.
- James, D. F. V., 1994. Change of polarization of light beams on propagation in free space. *Journal of the Optical Society of America A*, **11**, 1641.
- Jouttenus, T., Setälä, T., Kaivola, M., and Friberg, A. T., 2005. Connection between electric and magnetic coherence in free electromagnetic fields. *Physical Review E*, **72**, 046611.
- Kong, J. A., 2005. *Electromagnetic Wave Theory*. Cambridge, MA: EMW.
- Luis, A., 2007. Degree of coherence for vectorial electromagnetic fields as the distance between correlation matrices. *Journal of the Optical Society of America. A*, **24**, 1063–1068.
- Mandel, L., and Wolf, E., 1965. Coherence properties of optical fields. *Reviews of Modern Physics*, **37**, 231.
- Mandel, L., and Wolf, E., 1995. *Optical Coherence and Quantum Optics*. Cambridge, MA: Cambridge University Press.
- Setälä, T., Tervo, J., and Friberg, A. T., 2004. Complete electromagnetic coherence in the space–frequency domain. *Optics Letters*, **29**, 328.
- Tervo, J., Setälä, T., and Friberg, A. T., 2003. Degree of coherence for electromagnetic fields. *Optics Express*, **11**, 1137.
- Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley.
- Waxman, E., 2003. New direction for gamma-rays. *Nature*, **423**, 388.
- Wolf, E., 2003. Correlation-induced changes in the degree of polarization, the degree of coherence, and the spectrum of random electromagnetic beams on propagation. *Optics Letters*, **28**, 1078.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)

RADIATION, SOLAR AND LUNAR

David M. Le Vine

Code 615, Cryospheric Sciences Branch, NASA/Goddard Space Flight Center, Greenbelt, MD, USA

Definition

Solar Radiation: electromagnetic radiation emanating from the Sun with emphasis on microwaves.

Lunar Radiation: electromagnetic radiation from the Moon with emphasis on thermal emission in the microwave portion of the spectrum.

Solar radiation

Introduction

The Sun is a relatively typical star located about two thirds of the galactic radius from the center of the spiral (Aarons, 1965; Chap. 1). To a first approximation, the Sun has an effective temperature of about 5,800 K. This is the temperature of a thermal source with the same total flux (3.9×10^{26} W). At this temperature, the radiation peak is in the visible portion of the spectrum ($\lambda = 470$ nm).

However, the Sun is a dynamic structure that does not meet the conditions for thermal equilibrium, and there are considerable departures from this behavior, for example, in the UV (Aarons, 1965; Chap. 1) and in the microwave portion of the spectrum at wavelengths greater than about 1 cm (Kraus, 1966; Chap. 8.7). The level of radiation in the microwave portion of the spectrum follows the 11 year solar cycle associated with the number of sunspots and can reach extreme levels for brief periods associated with solar flares. For example, at 1.4 GHz, the effective temperature can range from 5×10^5 at solar minimum to more than 10^6 K near solar maximum, and the radiation can reach even higher levels during brief periods associated with solar flares (short-lived events associated with sudden brightening of a limited region on the solar disk). The discussion below focuses on the behavior in the microwave portion of the spectrum.

Terminology

Radiation from the Sun is commonly described in terms of “solar flux,” F_λ . The solar flux is the power density (W/m^2) incident at a given frequency (Hertz) integrated over the solar disk. Since the solar disk is small and assuming a mean value for the brightness ($\text{Wm}^{-2} \text{Hz}^{-1} \text{sr}^{-1}$), the solar flux is approximately

$$F_\lambda = B_{\text{sun}} \Omega_{\text{sun}} \quad (1)$$

where B_{sun} is the mean brightness and Ω_{sun} is the solid angle (steradians) of the Sun as seen from the Earth: $\Omega_{\text{sun}} = 2\pi[1 - \cos(\theta_{\text{sun}})] = 8.22 \times 10^{-5}$ sr. In this expression, θ_{sun} is the angular radius of the Sun as viewed from the Earth. The angular radius depends on wavelength and the physics generating the radiation. At 21 cm (1.4 GHz), $\theta_{\text{sun}} \approx 0.293^\circ$ which is about 10 % greater than the optical angular radius (Aarons, 1965; Chap. 3). The integral over space and frequency (called “luminosity”) is about 3.9×10^{16} W but this is dominated by radiation in the optical and nearby portions of the spectrum.

Values of F_λ expressed in solar flux units, sfu ($1 \text{ sfu} = 10^{-22} \text{ Wm}^{-2} \text{Hz}^{-1}$), are available at various radio wavelengths from ground stations around the world that monitor the Sun daily. Among these are the observatories in the Radio Solar Telescope Network (RSTN) operated by the US Air Weather Service (Sagamore Hill, 42 °N, 71 °W; Learmonth, Australia 22 °S, 114 E; Palehua, Hawaii 21 °N, 158 °W; and San Vito, Italy 41 °N, 18 E). Data from the RSTN can be obtained through the National Geophysical Data Center at Boulder, Colorado (NGDC website).

The radiation in solar flux units can be converted into an equivalent blackbody temperature by assuming that the Sun is a uniform thermal source (blackbody) and using the Rayleigh-Jeans approximation to the Planck radiation law for radio frequencies (Kraus, 1966 Chap. 3.9; Ulaby et al., 1981 Chap. 4.3). In this case, the relationship between the equivalent temperature, T_{sun} , and the solar flux, F_λ , is

$$T_{\text{sun}} = [\lambda^2/2k]B_{\text{sun}} \approx .00044\lambda^2F_\lambda \quad (2)$$

where k is the Boltzmann constant, λ is in meters, F_λ is in solar flux units, and T_{sun} is in Kelvin.

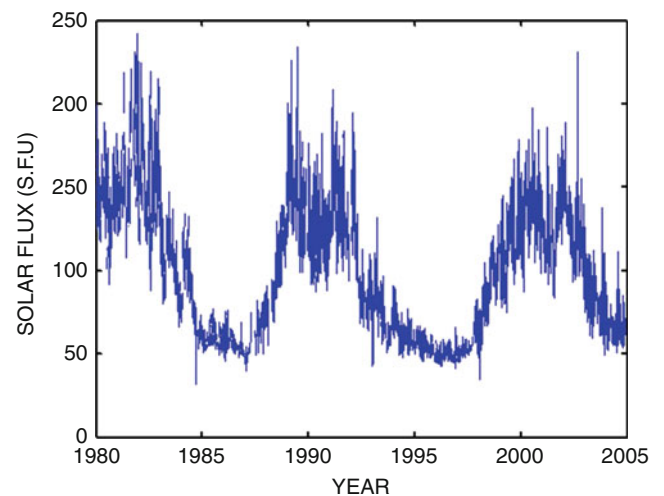
Characteristics

Radiation from the Sun follows the 11 year sunspot cycle, being relatively low during solar minimum (quiet Sun) and large and variable during solar maximum. It is convenient to separate radio emission from the Sun into two categories on the basis of their characteristic scale of temporal variation (Kraus, 1966, Chap. 8.7)

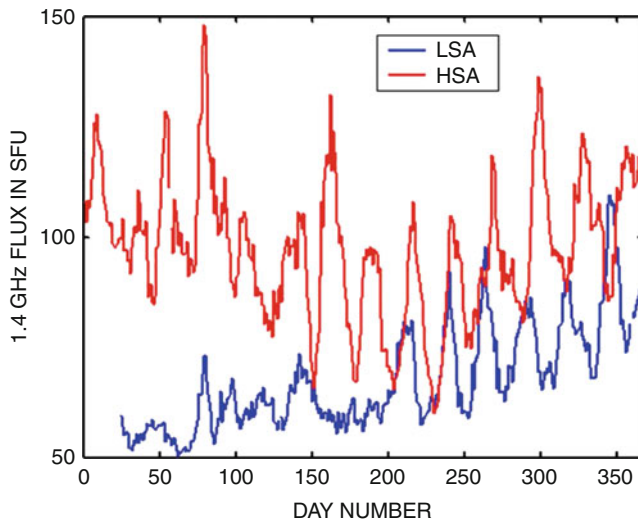
$$T_{\text{sun}}(t) = T_S(t) + T_R(t) \quad (3)$$

T_S is a slowly varying component (timescale of hours to days) which consists of two parts: (a) a relatively stable background level associated with the quiet Sun (solar minimum) that is largely due to thermal radiation from the hot plasma in the atmosphere of the Sun and (b) superimposed on this is a varying component that correlates with number of sunspots. The latter component changes irregularly with an amplitude that varies with the 11 year solar cycle (small at solar minima and large at solar maxima). Figure 1 shows the T_S component of the solar flux at 1.4 GHz measured at the Sagamore Hill solar observatory (these data are available from the NGDC website). The daily value of flux measured at local noon is plotted in sfu for the period 1980–2004 which includes the peak of three solar cycles. Figure 2 displays data for a period of 1 year with higher temporal resolution. The blue curve is for a minimum in the solar cycle (1986) and the red curve (1989) is near the maximum of the same solar cycle.

T_R is a rapidly varying component (seconds to minutes in duration) and is associated with flares and other transient activity on the Sun. Solar flares occur due to a sudden release of magnetic energy in the solar



Radiation, Solar and Lunar, Figure 1 Solar flux at 1.4 GHz at noon from Sagamore Hill (Le Vine et al., 2005).



Radiation, Solar and Lunar, Figure 2 Solar flux at 1.4 GHz at noon. High solar activity (1989, red) and low solar activity (1986, blue) (Data are from Sagamore Hill; Le Vine et al., 2005).

atmosphere and their frequency varies with the 11 year solar cycle. During periods of low solar activity, few solar flares are detected but there can be thousands during periods of high solar activity. Not all flares are associated with an enhancement of radiation at microwave frequencies. When such enhancements occur, they are called solar microwave bursts (SMB). Over the range 1–18 GHz, the spectrum of this radiation has a characteristic shape with a double peak at the low frequency end (1–2 GHz) and then a gradual decrease with frequency. A worldwide network of observing stations reports the SMB. Stations in the Radio Solar Telescope Network (RSTN) observe at 1 s intervals at eight frequencies in the microwave range 1–15 GHz. Bursts are reported at these stations only when flux exceeds 50 sfu. The Owens Valley Solar Array (OVSA) records solar flares with lower thresholds (OVSA website). An example of the distribution of peak amplitude and duration at 1.4 GHz is shown in Figure 3 (Le Vine et al., 2005).

Lunar radiation

The Moon is a source of both thermal emission and reflected solar radiation. At radio frequencies (e.g., microwave portion of the spectrum and longer wavelengths), the reflected solar radio emission is negligibly small compared to the thermal emission (Mayer, 1964). This is in contrast to optical frequencies, where reflected solar radiation is the dominant contribution (Evans, 1961).

Thermal emission from the Moon at radio frequencies was first reported by Dicke and Beringer (1946) and by Piddington and Minnett (1949). The source is believed to be the heating of lunar regolith due to solar radiation. Piddington and Minnett (1949) demonstrated a

dependence on lunar phase at 1.25 cm, but the dependence on phase decreases with increasing wavelength and the radiation is nearly independent of lunar phase at 21 cm (Hagfors, 1970).

The effective temperature of the Moon for thermal radiation depends on frequency. In the infrared, the temperature depends on the solar illumination and varies from about 400 K when the Moon is illuminated by the Sun (day time) to about 100 K at night (Hagfors, 1970). In the microwave portion of the spectrum, the dependence on lunar phase decreases from a change on the order of ± 115 K at 0.1 cm to negligible variation at 21 cm (Hagfors, 1970). The effective temperature for thermal emission in this portion of the spectrum is somewhat uncertain, ranging from about 200 to 275 K depending on the investigator, but the value is relatively independent of frequency (Hagfors, 1970). A median value for all investigations in the microwave portion of the spectrum is about 220 K (Hagfors, 1970, Table 3).

Impact of the sun and moon on remote sensing

Radiation from the Sun and Moon can impact passive remote sensing when looking down toward the surface of the Earth in several ways: (a) Radiation can enter the antenna via the “direct ray” along the line of sight from source to sensor and (b) Radiation can be reflected or scattered from the surface to the sensor. In the case of radiation from the Sun, the signal scattered from the ocean is often called “glitter.”

Direct ray

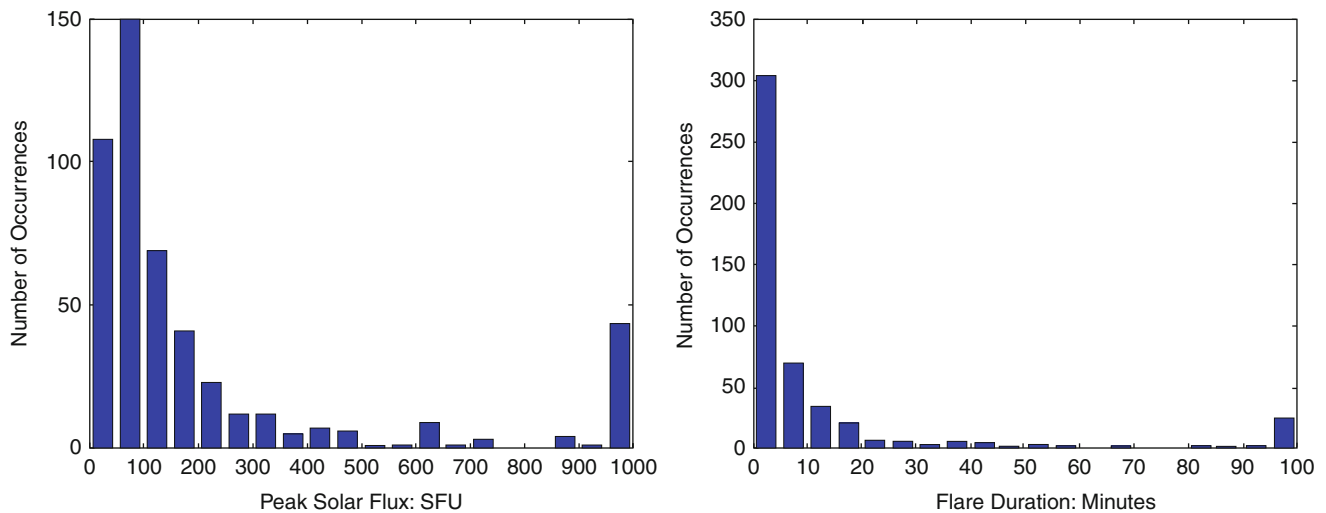
The contribution to the antenna temperature, T_A , can be obtained by integrating the incident radiation weighted by the antenna pattern over the extent of the source. Using the Rayleigh-Jeans approximation to express the radiation in terms of equivalent temperature, one obtains (Kraus, 1966, Chap. 3.18; Ulaby et al., 1981, Chaps. 4.4–6)

$$T_A = (1/4\pi) \int T_{\text{source}}(\Omega) G(\Omega) d\Omega \quad (4)$$

where Ω is solid angle and G is the antenna gain (Ulaby et al., 1981, Chap. 3.2). If T_{source} is constant and the source is small compared to the main beam of the antenna, then to a good first approximation,

$$T_A = (G/4\pi) T_{\text{source}} \Omega_{\text{source}} \quad (5)$$

As viewed from the Earth, the solid angles of the Sun and Moon (Ω_{source}) are very similar. However, T_{sun} is orders of magnitude larger than T_{moon} (e.g., $T_{\text{sun}} \approx 500,000$ K even for the quiet conditions, whereas $T_{\text{moon}} \approx 220$ K). The contribution of the direct rays from the Sun can be significant even for good antennas with low sidelobes in the direction of the Sun, but the direct ray from the Moon is almost always negligible.



Radiation, Solar and Lunar, Figure 3 Histograms showing the distribution of peak amplitude (*left*) and duration (*right*) for solar microwave bursts at 1.4 GHz during a period (1989) of high solar activity (Le Vine, et al., 2005).

Reflected ray

The effect of the reflected ray can be computed in the same manner except that the radiation is modified by reflection at the surface and the reflected ray encounters the antenna from a different direction. Using Eq. 5, one obtains

$$T_A = (G/4\pi)T_{\text{source}}\Omega_{\text{source}}R \quad (6)$$

R is a coefficient, called “reflectivity” to account for the reflection at the surface. In the case of a smooth surface, R is the magnitude squared of the Fresnel reflection coefficient. But, depending on the location of the observer, it may have to be modified to include effects of curvature of the Earth’s surface. Curvature will cause the reflected waves to diverge. For distant sources like the Sun and Moon, this can be taken into account by multiplying the reflectivity of a flat surface by a “divergence coefficient,” D (Beckmann and Spizzichino, 1963). In the case of a sphere, one finds that (Dinnat et al., 2009)

$$1/D \approx [1 + (2h/R_e \cos \gamma)][1 + (2h/R_e)] \quad (7)$$

where h is the distance from the reflection point to the spacecraft (i.e., slant range), R_e is the mean radius of the Earth (6,371 km), and γ is the angle between the ray from the source and the normal to the surface at the point of reflection. R is on the order of 0.7 for the ocean surface (no roughness) and 0.3 for land. For a satellite in low Earth orbit (about 800 km), D is on the order of 0.6, except for small values of $\cos(\gamma)$.

As in the case of the direct ray, reflected radiation from the Sun is much stronger than radiation from the Moon. The contribution due to the Moon is almost always negligible. However, in the case of a sensor looking down at the surface of the Earth, it is possible for radiation from the Moon to be reflected into the main beam of the antenna.

In this case, changes in T_A on the order of 1 K are possible (Dinnat et al., 2009).

The discussion above does not include the impact of roughness (e.g., waves on the ocean). See section 02.01.03 for information on scattering from rough surfaces.

Summary

To a first approximation, the Sun is a thermal source with an effective temperature of about 5,800 K. However, the radiation varies with the 11 year solar cycle and radio frequencies can greatly exceed this level during periods of high solar activity. The Moon is a source of both thermal emission and reflected solar radiation. In the microwave portion of the spectrum and longer wavelengths, the reflected solar radio emission is negligibly small compared to the thermal emission. At optical wavelengths, the reflected solar radiation is the dominant contribution.

Bibliography

- Aarons, J., 1965. *Solar System Radio Astronomy*. New York: Plenum.
- Beckmann, P., and Spizzichino, A., 1963. *The Scattering of Electromagnetic Waves from Rough Surfaces*. New York: Pergamon.
- Dicke, R. H., and Beringer, R., 1946. Microwave radiation from the sun and the moon. *The Astrophysical Journal*, **103**, 375.
- Dinnat, E. P., Abraham, S., Le Vine, D. M., de Matthaeis, P., and Jacob, D., 2009. Effect of emission from the moon on remote sensing of sea surface salinity: an example with the aquarius radiometer. *IEEE Transactions on Geoscience and Remote Sensing Letters*, **6**(2), 239–243.
- Evans, J. V., 1961. Radio echo studies of the moon. In Kopal, Z. (ed.), *Physics and Astronomy of the Moon*. New York: Academic, pp. 429–479.
- Hagfors, T., 1970. Remote probing of the moon by infrared and microwave emissions and by radar. *Radio Science*, **5**(2), 189–227.

- Kraus, J. D., 1966. *Radio Astronomy*. New York: McGraw-Hill.
- Le Vine, D. M. S., Wentz, A. F., and Lagerloef, G. S. E., 2005. Impact of the sun on remote sensing of sea surface salinity from space. In *Proceedings of International Geoscience and Remote Sensing Symposium (IGARSS)*, Seoul Korea.
- Mayer, C. H., 1964. Thermal radio radiation from the moon and planets. *IEEE Transactions on Antennas Propagation*, **AP-12**(7), 902–913.
- National Geophysical Data Center. <http://www.ngdc.noaa.gov/stp/SOLAR/ftp/solarradio.html>.
- Owens Valley Solar Array. <http://www.ovsa.njit.edu/>.
- Piddington, J. H., and Minnett, H. C., 1949. Microwave thermal radiation from the moon. *Australian Journal of Scientific Research*, **A2**, 63–77.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing*. Reading, MA: Addison-Wesley, Vol. I.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)
[Optical/Infrared, Radiative Transfer](#)
[Radiation, Electromagnetic](#)

RADIATION, VOLUME SCATTERING

Leung Tsang¹ and Kung-Hau Ding²

¹Paul Allen Center, Department of Electrical Engineering, University of Washington, Seattle, WA, USA

²Air Force Research Laboratory, Wright-Patterson AFB, Dayton, OH, USA

Definition

Volume scattering. Multiple scattering events taking place when electromagnetic wave inside a medium containing scatterers with discrete dielectric properties.

Radiative transfer. The transport of electromagnetic radiation power, scattered or absorbed, while propagating through a medium containing inhomogeneities.

Specific intensity. The instantaneous radiation power per unit area, per unit frequency interval at a specific frequency, at a position inside the medium, and in a specific direction per unit solid angle.

Independent scattering. Assumption of particles being far apart from each other, separation much greater than the incident wavelength, such that no correlation of scattered fields from different scatterers.

Introduction

The volumetric scattering effects play an important role in radar remote sensing of geophysical media (Tsang et al., 1985; Tsang and Kong, 2001). The volume scattering contributes to the bistatic scattering in the active radar remote sensing. In the passive remote sensing, the volume scattering redistributes the thermal emission. Electromagnetic waves have amplitudes and phases. The phase depends on the distance of propagation. However, in geophysical media, the scatterers are randomly positioned, so that the observed wave in bistatic directions

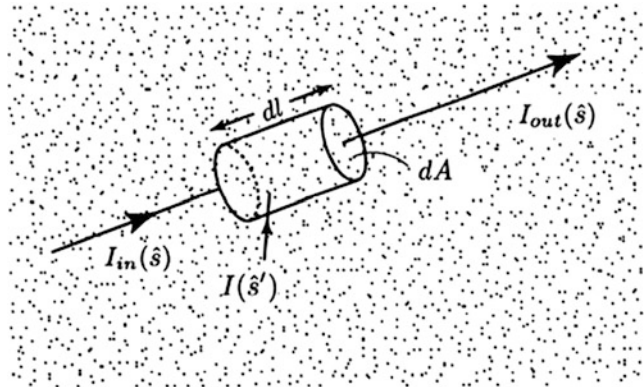
usually has random phase as the wave propagates many wavelengths of distances. Such a wave is called the “incoherent wave.” Although the final observed wave is incoherent, the wave interactions among the scatterers can have coherent effects. For example, in microwave scattering by ice grains in snow, there are many ice grains within the distance of a wavelength. Thus, the wave interactions among particles in the neighborhood of each other are coherent. Another example is the scattering by vegetation. Because of the partial ordering of the trunks, branches, and leaves, partial coherence effects are also exhibited in the phase difference between the horizontal and vertical polarizations.

Radiative transfer theory

Consider a medium consisting of scatterers that are randomly distributed, as shown in Figure 1. Inside the medium, the differential change of the specific intensity $\bar{I}(\bar{r}, \hat{s})$ at the position $\bar{r}$ and in the direction $\hat{s}$ obeys the radiative transfer equation:

$$\frac{d\bar{I}(\bar{r}, \hat{s})}{ds} = -\kappa_e \bar{I}(\bar{r}, \hat{s}) + \kappa_a T \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \int d\hat{s}' \bar{P}(\hat{s}, \hat{s}') \cdot \bar{I}(\bar{r}, \hat{s}') \quad (1)$$

in which κ_e is the extinction coefficient that is a summation of absorption and scattering and κ_a is the absorption coefficient. The term $\kappa_e \bar{I}(\bar{r}, \hat{s})$ in the radiative transfer equation represents the decrease of the specific intensity $\bar{I}$ due to extinction, while the term $\kappa_a T$ represents an increase of specific intensity due to microwave emission that is proportional to the temperature of



Radiation, Volume Scattering, Figure 1 A random medium consists of a large number of scatterers. Due to scattering, the distribution of the specific intensity $\bar{I}$ is over all directions $\hat{s}$ and all positions $\bar{r}$. The differential change in the specific intensity as it passes through a small volume is along the direction of $\hat{s}$.

medium T . The term $\int d\hat{s}' \bar{\bar{P}}(\hat{s}, \hat{s}') \cdot \bar{I}(\hat{r}, \hat{s}')$ inside the integral of the radiative transfer equation represents an increase of the specific intensity due to the scattering of $\bar{I}$ from the direction $\hat{s}'$ into the direction $\hat{s}$. This increase is proportional to the specific intensity $\bar{I}(\hat{r}, \hat{s}')$ in the direction of $\hat{s}'$. The quantity $\bar{\bar{P}}(\hat{s}, \hat{s}')$ is the phase matrix that is related to the scattering properties of the particles and describes the coupling of specific intensities from the direction $\hat{s}'$ to the direction $\hat{s}$. The vector notation of the specific intensity $\bar{I}$ is because it has four components. The first two components are the vertically polarized intensity and the horizontally polarized intensity. The third and fourth components are respectively the real part and the imaginary part of the correlation between the vertical polarization and the horizontal polarization.

Solution of radiative transfer equation and relations to remote sensing signatures

Remote sensing consists of active remote sensing using radars and passive remote sensing using radiometers. In the active remote sensing, as illustrated in Figure 2 for a snow layer, we can neglect the emission term in the radiative transfer Eq. 1. An incident wave with a specific intensity $\bar{I}_0$ is impinging in the $\hat{s}_0$ direction. We then solve the radiative transfer equation subject to the boundary conditions at the boundary $z = 0$ and at the boundary $z = -d$. The bistatic scattering coefficient will be proportional to the resultant-specific intensity $\bar{I}(z = 0, \hat{s})$ at the boundary $z = 0$. The backscattering coefficient is for the specific intensity $\bar{I}(z = 0, \hat{s} = -\hat{s}_0)$ at the direction $\hat{s} = -\hat{s}_0$.

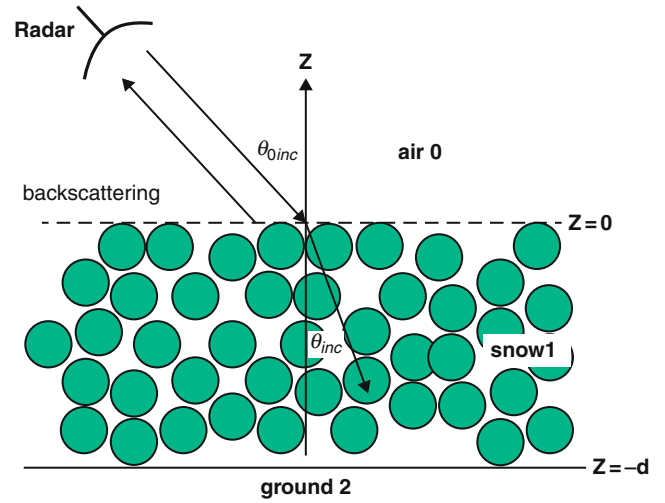
In the passive remote sensing, as depicted in Figure 3 for a snow medium, the boundary condition of the radiative transfer equation for a smooth surface at $z = 0$ is given by

$$\bar{I}(\hat{s}_d) = \bar{\bar{R}}_{10} \cdot \bar{I}(\hat{s}_u) \quad (2)$$

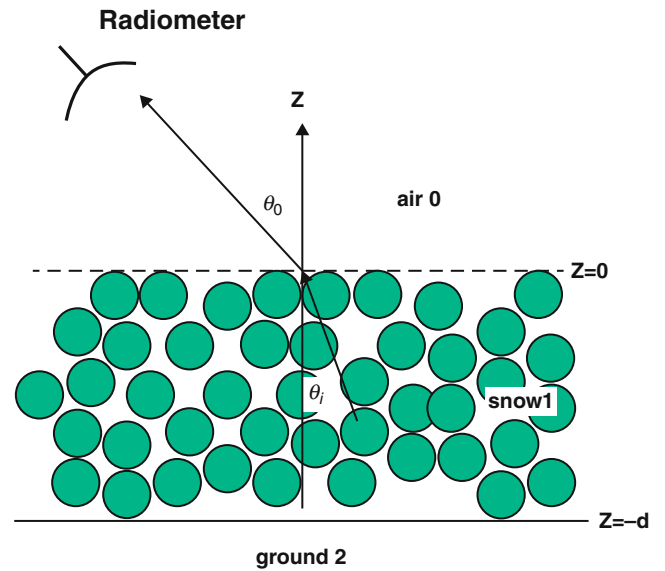
where $\bar{\bar{R}}_{10}$ is the Fresnel reflectivity for a smooth surface between medium 1 and medium 0, $\hat{s}_u$ denotes the upward direction, and $\hat{s}_d$ is for the downward direction corresponding to the specular reflection of $\hat{s}_u$. This boundary condition is saying that the downward intensity at the interface of media 0 and 1 is equal to the upward intensity in the medium 1 reflected by the boundary between the media 1 and 0. At the boundary $z = -d$, the interface between medium 1 and medium 2, the boundary condition is

$$\bar{I}(\hat{s}_u) = \bar{\bar{R}}_{12} \cdot \bar{I}(\hat{s}_d) + \bar{\bar{T}}_{12} \cdot \bar{T}_2 \quad (3)$$

in which $\bar{\bar{R}}_{12}$ and $\bar{\bar{T}}_{12}$ are respectively the Fresnel reflectivity and transmissivity for a smooth surface between medium 1 and medium 2 and $\bar{T}_2$ is the temperature of medium 2. This boundary condition says that the upward intensity in the medium 1 is equal to the sum of the downward intensity in the medium 1 reflected by the boundary between media 1 and 2 and the emissions from the medium 2.



Radiation, Volume Scattering, Figure 2 Active microwave remote sensing: microwave from radar to snow is scattered and attenuated by snow grains. Backscattered microwaves received by radar give the backscattering coefficient.



Radiation, Volume Scattering, Figure 3 Passive microwave remote sensing: ground emission and snow emission are scattered and attenuated by snow grains before received by radiometer as brightness temperature.

The radiative transfer equations are to be solved subject to the boundary conditions. The brightness temperature $\bar{T}_B$ as measured by a radiometer can be expressed as

$$\bar{T}_B(\hat{s}_{0u}) = \bar{\bar{T}}_{10} \cdot \bar{I}(z = 0, \hat{s}_u) \quad (4)$$

where the directions $\hat{s}_{0u}$ and $\hat{s}_u$ are related by the Snell's law and $\bar{\bar{T}}_{10}$ is the transmissivity for a smooth surface between medium 1 and medium 0.

Solution techniques of radiative transfer equations

There are several techniques of solving the radiative transfer equations (Tsang et al., 1985). One technique consists of an iterative approach by assuming that the phase matrix is small. Then one can solve the radiative transfer equation by first neglecting the phase matrix term. The solution is then obtained by solving a first-order ordinary differential equation. The solution is the zeroth-order solution. Then one adds the phase matrix term back and uses the zeroth-order solution to obtain the first-order solution. One can proceed further to obtain the higher-order solutions.

A second method is the eigen-quadrature approach. It consists of discretizing the propagation direction of the specific intensity, for example, into 16 directions. Then the integral term $\int d\hat{s}' \bar{P}(\hat{s}, \hat{s}') \cdot \bar{I}(\hat{r}, \hat{s}')$ becomes a product of a matrix and a column vector. The radiative transfer equation becomes a system of ordinary differential equations which are solved by the summation of the particular solution and the eigensystem for the homogeneous solution. The coefficients are obtained by matching boundary conditions. The solution by this eigen-quadrature method includes the full multiple scattering effects within the radiative transfer theory.

A third approach is by using the method of invariant embedding or by the doubling of layers. The essence of this approach is based on recognizing that the bistatic scattering properties of a layer of thickness much less than the optical thickness can be expressed in terms of the bistatic scattering properties of a layer of half of its thickness. By using this property, the solution of a layer of finite thickness can be obtained by starting with a thin layer and then continuing the doubling of layer thickness until the desired thickness is obtained.

Independent scattering: calculation of scattering and absorption coefficients and phase matrix

To relate to the physical properties of the medium, it is required to express the scattering coefficient, absorption coefficient, and phase matrix in terms of the physical properties of the constituent particles, such as fractional volume, particle sizes, shapes, orientations, and permittivities (Tsang et al., 1985, 2001). The classical method is based on the independent scattering. The independent scattering means that the scattering of a conglomeration of particles is equal to the addition of individual particle scattering properties. Under the independent scattering assumption, the scattering coefficient κ_s of a medium with identical particles is equal to

$$\kappa_s = n_0 \sigma_s \quad (5)$$

where n_0 is the number of particles per unit volume and σ_s is the scattering cross section of a single particle. Similarly, the absorption coefficient κ_a is given by

$$\kappa_a = n_0 \sigma_a \quad (6)$$

where σ_a is the absorption cross section of a single particle. The phase matrix of a medium with identical particles is given by

$$\bar{P}(\hat{s}, \hat{s}') = n_0 \bar{\sigma}_{bs}(\hat{s}, \hat{s}') \quad (7)$$

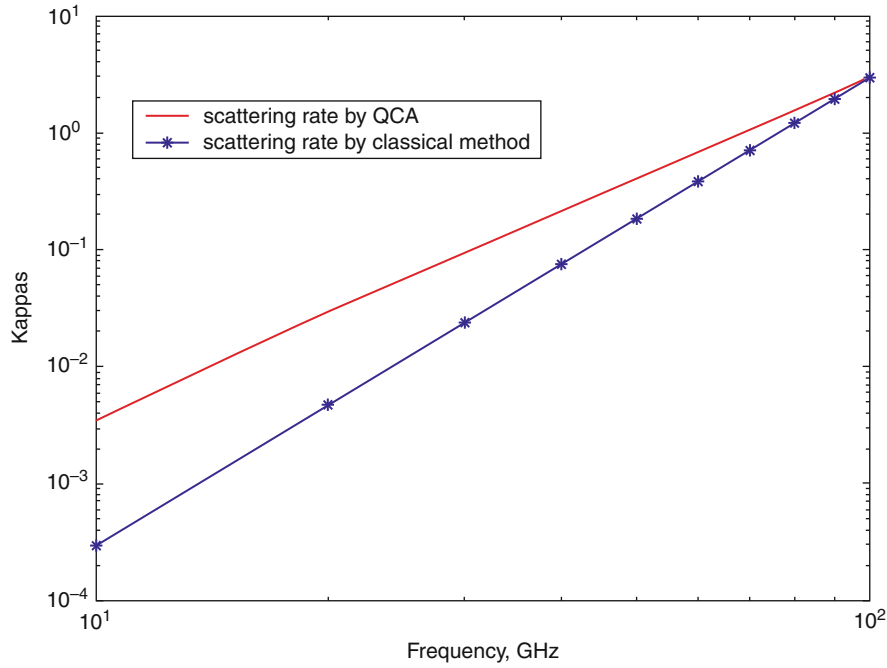
in which $\bar{\sigma}_{bs}(\hat{s}, \hat{s}')$ is the bistatic scattering cross section of a single particle.

Maxwell equations and coherent effects

Electromagnetic wave propagation and scattering in natural or man-made media are governed exactly by Maxwell equations. Thus, one can solve Maxwell equations for a collection of particles to calculate various electromagnetic properties, such as the scattering, absorption, and extinction coefficients. However, in Maxwell equations, electromagnetic fields have amplitudes and phases. Such a solution will exhibit interference effects of scattering by different particles. Therefore, the solutions of Maxwell equations have to be averaged over realizations or samples of collections of particles. If the medium is statistically homogeneous and the concentration of particles is sparse, the solutions of Maxwell equations, when averaged over realizations, will agree with those of independent scattering. The independent scattering approach is usually employed to study the scattering of atmospheric particles, for example, rain, fog, and cloud.

Scattering by dense media

In dense media, the constituent scatterers are closely packed. In densely packed media, the particles do not scatter independently. The correlation between particles has to be taken into account. A dense medium theory takes into account the collective and dependent scattering of particles. For example, snow is a dense medium because the ice grains in snow are densely packed with the volume fraction of ice ranging from 10 % to 50 %. In the dense medium theory, the analytic method consists of using the quasicrystalline approximation (QCA) to calculate the scattering properties of the particles (Tsang et al., 1985; Tsang and Kong, 2001). The scattering signatures of dense media depend on the pair distribution functions of particles which describe the relative positions of the particles. The phase matrices of dense media depend on the Fourier transform of the pair distribution functions. The dense medium approach gives scattering results distinctly different from those of independent scattering. For example, for a collection of particles that are much smaller in size than a wavelength, independent Rayleigh scattering gives the scattering coefficient that depends on frequency to the fourth power. However, from the theory of quasicrystalline approximation (QCA), the aggregated property of densely packed particles gives weaker frequency dependence than the fourth power. For example, consider a dense medium with spherical particles of diameter 1.2 mm, dielectric



Radiation, Volume Scattering, Figure 4 Scattering rate: quasicrystalline approximation (QCA) compared with classical scattering theory. The particle size is 1.2 mm and the volume fraction is 20 %.

constant 3.2, and volume fraction 20 %. In Figure 4, we compare the scattering coefficients between QCA and the classical scattering theory as a function of frequency for this medium. It shows that the scattering coefficient of QCA has weaker frequency dependence than that of the classical theory (Chen et al., 2003; Tse et al., 2007).

Foldy-Lax multiple scattering equations

In addition to the analytic approach, one can also use the Numerical Maxwell Model of 3-Dimensional (NMM3D) simulations to study the scattering by dense media (Tse et al., 2007). In the NMM3D method, the positions of particles in a medium are randomly computer generated. The particles are also allowed to bond together to form aggregates. Maxwell equations are then solved numerically. A common approach of solving Maxwell equations in multiple scattering by many scatterers is through the use of Foldy-Lax multiple scattering equations (Tsang et al., 1985, 2001).

Consider a collection of N particles; the Foldy-Lax multiple scattering equations are of the following form:

$$\bar{E}_j^{ex} = \bar{E}^{inc} + \sum_{\ell=1, \ell \neq j}^N \bar{G}_0 \cdot \bar{T}_\ell \cdot \bar{E}_\ell^{ex} \quad (8)$$

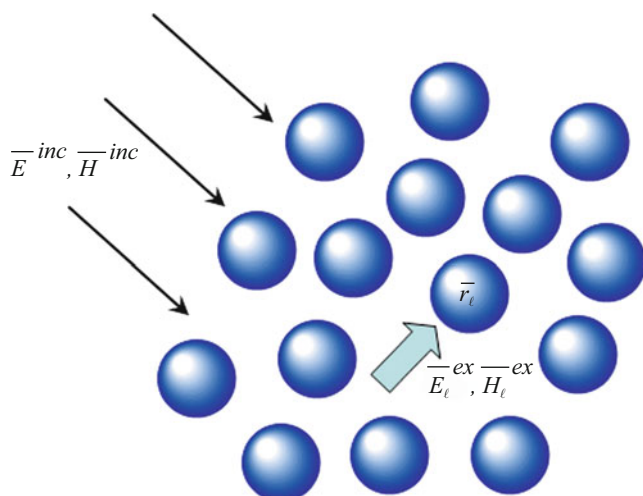
It says that the field exciting the particle j , $\bar{E}_j^{ex}$, is equal to the sum of the electric field of the incident wave, $\bar{E}^{inc}$,

and the scattered wave from all other particles excluding itself, as shown in Figure 5. In the Foldy-Lax scattering equation, $\bar{G}_0 \cdot \bar{T}_\ell \cdot \bar{E}_\ell^{ex}$ gives the scattered wave from the particle ℓ to the particle j , and the summation symbol $\sum_{\ell=1, \ell \neq j}^N$ represents the sum of such scattered waves. The transition matrix or T-matrix, $\bar{T}_\ell$, of particle ℓ acting on the exciting field, $\bar{E}_\ell^{ex}$, of the ℓ particle gives the scattered wave of the ℓ particle which travels to the j^{th} particle through the Green's function or the propagator, $\bar{G}_0$. Note that we have exciting fields on both the sides of the Foldy-Lax equations.

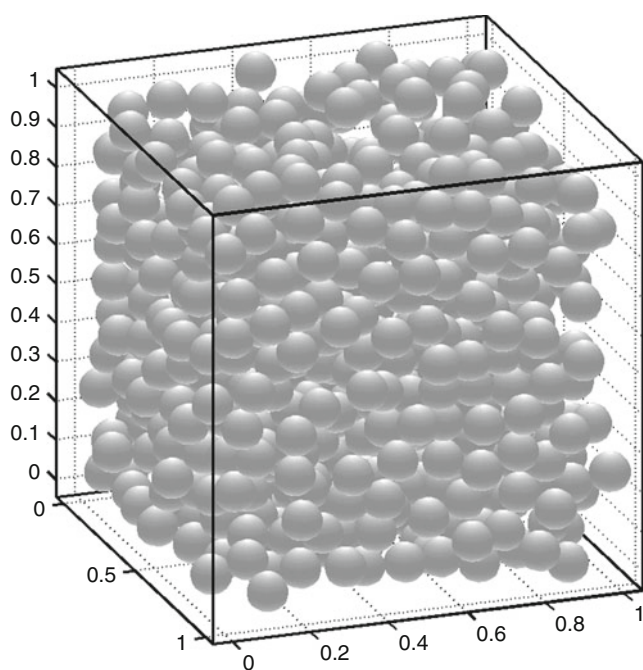
To solve the Foldy-Lax equations, vector wave expansions are applied and the Foldy-Lax scattering equations are cast into a matrix equation. The dimension of the matrix equation is equal to the number of particles, N , multiplied by the number of multipole terms used in the vector wave expansions. As a large number of particles are usually required to simulate the coherent and incoherent waves in dense media, the dimension of the resultant matrix equation is quite large. The Foldy-Lax equations can be solved using the iterative technique.

Results of QCA and NMM3D

For the case of dense media, the solutions of the Maxwell equation approach are consistent with QCA. However, the solution of Maxwell equations gives appreciable cross-polarization in the phase matrix even for a collection of spheres, a feature that is different from the independent scattering and from the quasicrystalline approximation.



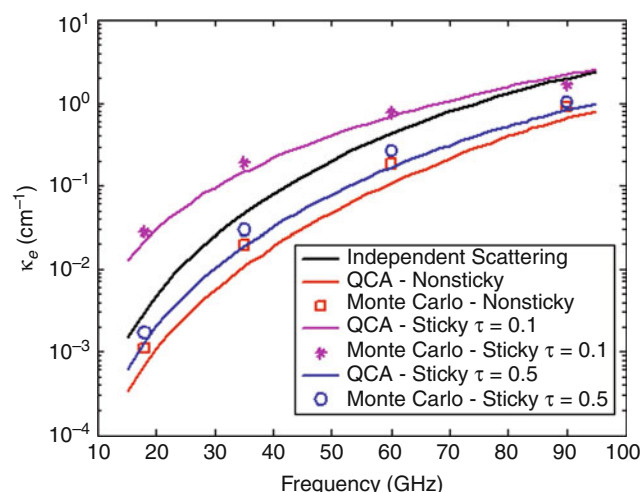
Radiation, Volume Scattering, Figure 5 A schematic diagram shows the incident electromagnetic fields and the exciting electromagnetic fields in a medium of randomly positioned scatterers.



Radiation, Volume Scattering, Figure 6 An example of computer-simulated dense medium with randomly distributed spherical particles.

In Figure 6, we show a collection of spherical particles densely packed within a cubic box based on the computer simulation.

In Figures 7 and 8, we compare the extinction coefficients for snow as a function of frequency. We show the results of independent scattering, QCA, Monte Carlo



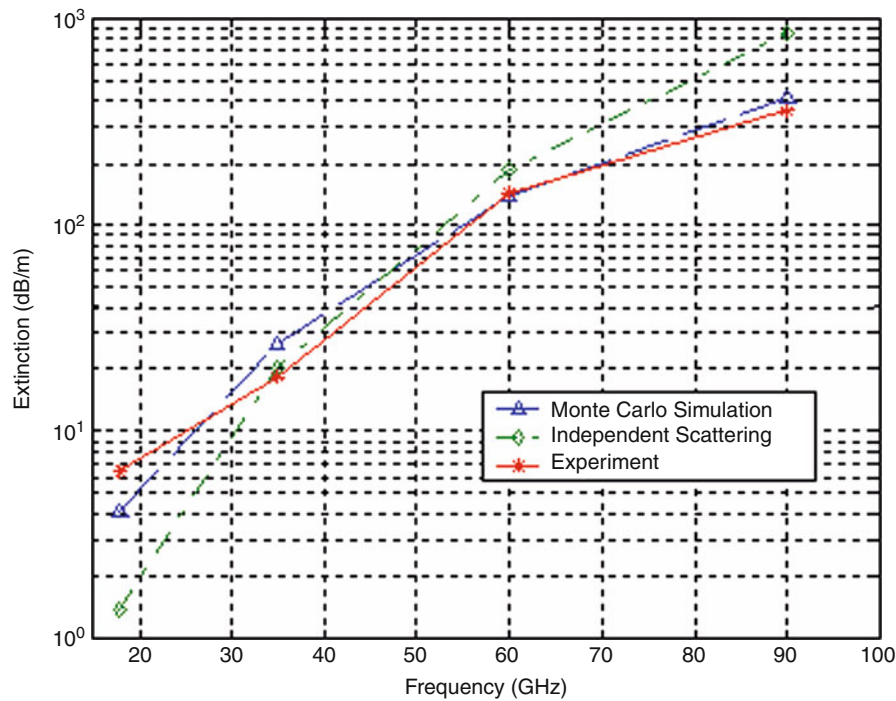
Radiation, Volume Scattering, Figure 7 Comparison of extinction coefficients (in $1/\text{cm}$) between the independent scattering, QCA, and Monte Carlo NMM3D results.

NMM3D, and those of experimental measurements. The results of NMM3D have the closest agreement with the experimental data that exhibit much weaker frequency dependence than that of independent scattering (Chen et al., 2003).

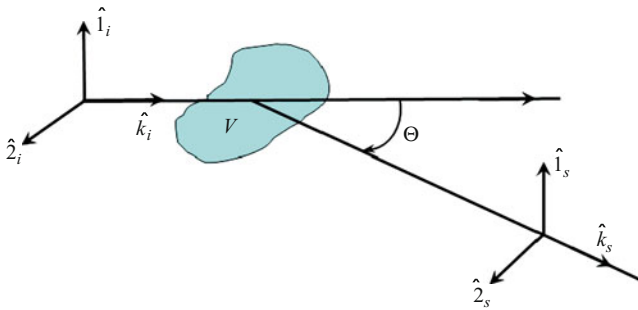
The phase matrices are best represented by the 1–2 frame (Tsang et al., 2000). The 1–2 frame is shown in Figure 9, in which $\hat{k}_i$ is the incident direction and $\hat{k}_s$ is the scattered direction. The polarization vectors $\hat{l}_i$ and $\hat{l}_s$, with $\hat{l}_i = \hat{l}_s$, are perpendicular to the plane formed by $\hat{k}_i$ and $\hat{k}_s$. The $\hat{2}_i$ direction is perpendicular to $\hat{k}_i$ and $\hat{l}_i$, while the $\hat{2}_s$ direction is perpendicular to $\hat{k}_s$ and $\hat{l}_s$.

In Figure 10, the phase matrix elements are plotted as a function of the angle Θ between the incident and scattered directions for two frequencies, 18 and 37 GHz. The medium considered is the same as in Figure 4. The results are shown and compared for the independent scattering and QCA approaches. The phase matrix component P_{11} is the scattering from the polarization $\hat{l}_i$ to the polarization $\hat{l}_s$, while P_{22} is the phase matrix component of scattering from the polarization $\hat{2}_i$ to the polarization $\hat{2}_s$. We note that for the same grain size, QCA predicts more forward scattering than the independent scattering (Tsang et al., 2007).

In Figure 11, we compare the brightness temperatures from QCA/DMRT predictions with experimental data of passive remote sensing of snow. Four channels of brightness temperatures are shown, with vertically and horizontally polarized brightness temperatures at 18.7 and 36.5 GHz. A multilayer snow model is used with varying snow density and grain size. We note that the brightness temperatures at 18.7 GHz are higher than those of the 36.5 GHz because the scattering at 36.5 GHz is stronger than that at 18 GHz. The brightness temperature for vertical polarization is higher than that of horizontal



Radiation, Volume Scattering, Figure 8 Comparison of extinction (in dB/m) between the independent scattering, Monte Carlo NMM3D results, and experimental data.



Radiation, Volume Scattering, Figure 9 Geometry for defining the 1-2 system based on the scattering plane. The scattering plane contains $\hat{k}_i$ and $\hat{k}_s$. The angle between $\hat{k}_i$ and $\hat{k}_s$ is Θ .

polarization because of stronger boundary reflection for horizontal polarization. The QCA/DMRT results are in good agreement with experimental observations (Liang et al., 2008).

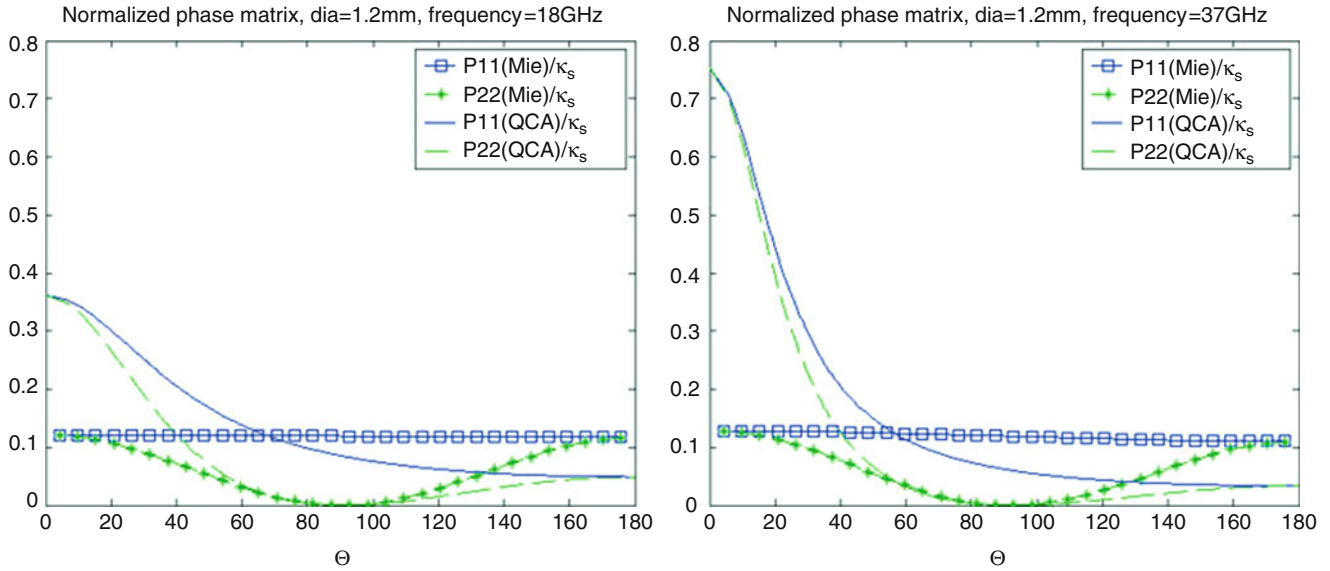
Scattering by vegetation

Various scattering models of vegetation have been developed for scattering of foliage, vegetation, and forests. We can divide them into three kinds of models: (1) empirical/physical model, (2) vector radiative transfer model, and (3) coherent approach and full-wave multiple scattering model. The empirical model is based on a simplified radiative transfer model with empirical fitting parameters

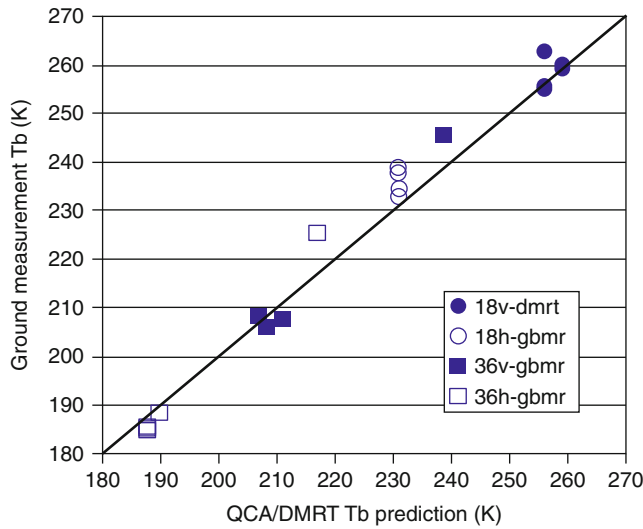
for various vegetation types. The vector radiative transfer approach is based on incoherent interactions of branches, leaves, and trunks (Tsang et al., 1985; Ulaby et al., 1990). The Michigan Microwave Canopy Scattering model (MIMICS) (McDonald et al., 1990) was commonly used based on a first-order vector radiative transfer theory. The coherent superposition approach takes into account the relative phase shifts between scattering from different branches and leaves. It accounts for clustering structures. A computer growth model of Lindenmayer system can be used to grow and characterize the clustering properties. The full-wave multiple scattering model takes into account the coherent wave interaction among branches and leaves. The advantages of empirical model are that they use the fewest parameters that have been tuned to experimental measurements. However, the tuning parameters can be site specific and vegetation type specific. Both vector radiative transfer model and coherent wave model are physically based. Numerical full-wave simulations can be used to validate analytical radiative transfer models. Physical models can also be used to compare with empirical models. In actual implementation of physical models, the number of physical parameters can be reduced by having fixed relations of the physical parameters that are tuned to vegetation types.

Empirical/simplified physical model

The usefulness of an empirical model is that the parameters can be related to experimental data. Models with the



Radiation, Volume Scattering, Figure 10 Comparison between QCA and the Mie scattering phase matrices. The particle size is 1.2 mm and the volume fraction is 20 %. The frequencies are 18 GHz (left) and 37 GHz (right).



Radiation, Volume Scattering, Figure 11 Comparison of brightness temperatures (in K) between ground observation data at CLPX and model QCA/DMRT simulations. The observation frequencies are 18.7 and 36.5 GHz.

fewest parameters are desirable. For vegetation volume scattering, the radar backscattering coefficient can be expressed as

$$\sigma = \sigma_v + \sigma_{vg} + \sigma_{gv} \quad (9)$$

in which the three terms are respectively the (1) direct volume scattering (σ_v), (2) the scattering by the ground which is subsequently scattered by the vegetation (σ_{vg}), and (3) the scattering by the vegetation that is

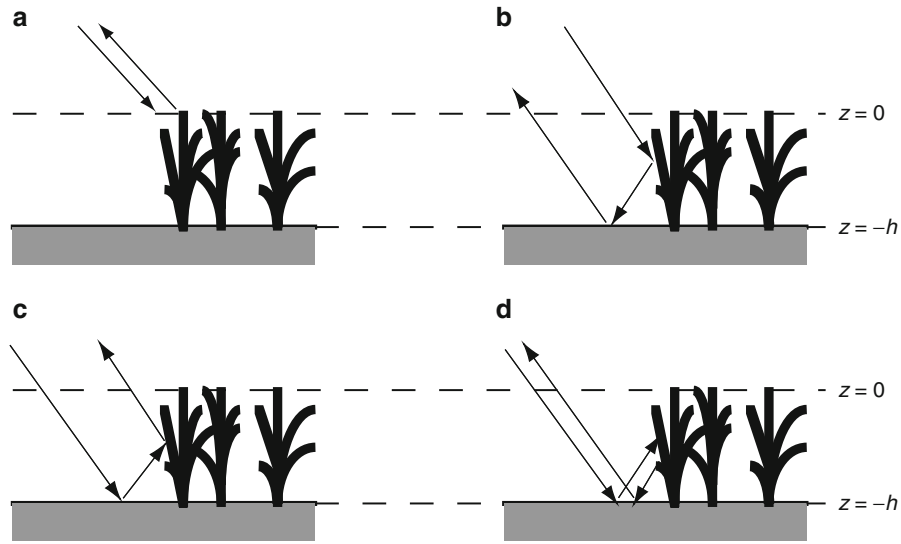
subsequently scattered by the ground (σ_{gv}). Empirical formulas can be established for each of the three terms using simplified radiative transfer models in a manner similar to that of the passive microwave remote sensing soil moisture algorithms. The empirical parameters can be in terms of optical thickness, τ , and the scattering albedo $\bar{\omega}$. Both parameters can be polarization dependent. The optical thickness can be related to the vegetation moisture content. The scattering albedo can be decomposed into like-polarization albedo and cross-polarization albedo. If we assume that the optical thickness can be retrieved from the vegetation moisture content that can be auxiliary information from NDVI, then the empirical model requires only two parameters, the like-polarization albedo and the cross-polarization albedo.

Vector radiative transfer theory

The vegetation canopy consists of leaves, branches, and trunks. In the theoretical scattering models, the leaves are usually assumed to be circular disks, the branches are treated as finite length cylinders, and the trunks are treated as vertical cylinders. Thus, based on the independent scattering assumption, the vector radiative transfer theory can be constructed for the vegetation scattering (Tsang et al., 1985; Ulaby et al., 1990):

$$\frac{d\bar{I}(\bar{r}, \hat{s})}{ds} = -\bar{\kappa}_e \cdot \bar{I}(\bar{r}, \hat{s}) + \bar{\kappa}_a T + \int d\hat{s}' \bar{P}(\hat{s}, \hat{s}') \cdot \bar{I}(\bar{r}, \hat{s}') \quad (10)$$

Because the circular disks and the cylinders are nonspherical objects with some orientation distributions, the extinction in the radiative transfer theory becomes



Radiation, Volume Scattering, Figure 12 Four major backscattering mechanisms in the vegetation canopy scattering model: (a) direct scattering; (b) single scattering followed by reflection; (c) reflection followed by single scattering; and (d) reflection followed by single scattering and further reflection.

a 4 by 4 extinction matrix, $\bar{\kappa}_e$, and the absorption is a 4 by 1 absorption vector, $\bar{\kappa}_a$. For active radar scattering applications, the temperature emission term can be ignored. The vector radiative transfer model can be decomposed into multilayers. The first layer consists of branches and leaves and the second layer consists of trunks. The vector radiative transfer equations are to be solved subject to boundary conditions at the vegetation and ground interface. The boundary conditions can be of rough surface boundary conditions or that of specular reflection (Tsang et al., 1985).

Phase matrices of branch and leaf

The components of phase matrices for branches and leaves have been computed based on approximate and exact solutions of Maxwell's equations (Tsang et al., 1992, 2000, 2001). However, care must be exercised in calculating the phase matrices in a manner that is consistent with energy conservation.

Approximate scattering amplitudes. The scattering models are that of disks to represent leaves and finite length cylinders to represent branches and trunks. Approximate solutions of these objects were constructed based on infinite cylinder approximation, physical optics, and small disks assumptions.

Exact scattering amplitudes based on BOR. The body of revolution (BOR) codes are based on the method of moments (MoM) and have been used to replace the approximate solutions (Tsang et al., 1992). It converts a 3D problem through rotational symmetry into a 2D solution of Maxwell equations. The CPU is presently

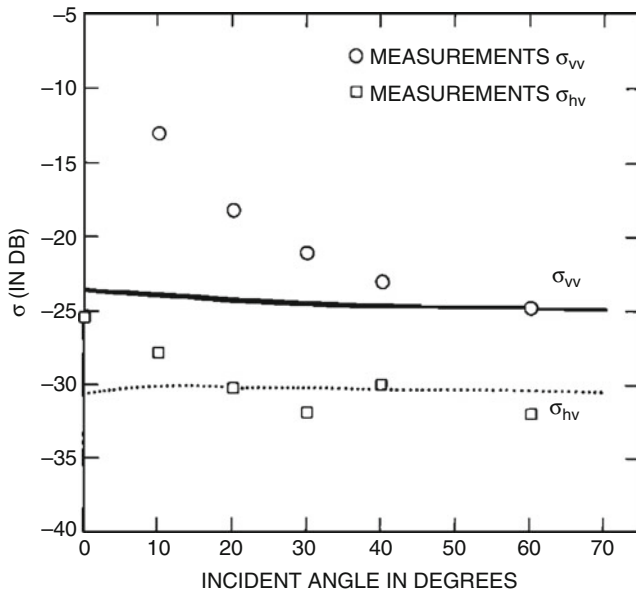
acceptable because it is a 2D problem and because of the advance of modern computers. The advantage of BOR is that it gives accurate forward scattering amplitudes which can be used in the computation of extinction matrix elements in the vector radiative transfer theory.

First-order solution

The first-order solution is valid when the optical thickness of the vegetation canopy is thin, such as equal to or less than 0.2. The first-order solution consists of four terms, as shown in Figure 12. They are (1) direct volume scattering, (2) scattering by the ground which is subsequently scattered by the vegetation, (3) scattering by the vegetation that is subsequently scattered by the ground, and (4) scattering by the ground, then by the vegetation, and then by the ground again. These are four distinct contributions for bistatic scattering. The radar scattering coefficient can be expressed as

$$\sigma = \sigma_v + \sigma_{vg} + \sigma_{gv} + \sigma_{gvg} \quad (11)$$

Within the radiative transfer theory, these four contributions are added incoherently for the bistatic scattering. For radar monostatic backscattering, contributions (2) and (3) are identical because of reciprocity. Furthermore, contributions (2) and (3) interfere constructively so that the field adds coherently and the intensity becomes four times in the exact backscattering direction. Thus, in the backscattering direction, the incoherent approach gives $\sigma = \sigma_v + 2\sigma_{vg} + \sigma_{gvg}$, while the coherent approach gives $\sigma = \sigma_v + 4\sigma_{vg} + \sigma_{gvg}$. Whether the factor should be "2" or "4" or somewhere between 2 and 4 depends on the angular width of the backscattering enhancement and the radar beam width. These issues are yet to be studied



Radiation, Volume Scattering, Figure 13 Comparison of theoretical results with backscattering measurements from soybean at 1.1 GHz. The parameters used in the theoretical model are ground permittivity $\epsilon_2 = 4\epsilon_0$; depth $d = 1$ m; fractional volume 0.3 %; particles permittivity $\epsilon_s = (30.66 + i1.7)\epsilon_0$; major axis $a = 1.5$ cm; minor axis $c = 0.02$ cm; and the orientation angles $\beta_2 = \pi/2$, $\gamma_1 = 0$, and $\gamma_2 = 0$.

(Tsang et al., 1985; Tsang and Kong, 2001). To illustrate the usefulness, we make a comparison of theoretical results in Figure 13 with backscattering measurements from soybeans (Dobson et al., 1977). The simulation uses a layer of randomly positioned and oriented elliptical particles.

Second-order radiative transfer theory

This theory consists of a second iteration of the vector radiative transfer theory and is useful when the optical thickness is between 0.2 and 0.8. It includes the effects of double volume scattering in vector radiative transfer theory (Tsang et al., 1992).

Coherent approach and full-wave multiple scattering

The vector radiative transfer approach assumes that the scatterers scatter incoherently so that the intensities add. For wave propagation and scattering, the fields are coherent with amplitude and phase. Thus, much work has been done in using a coherent approach and a full-wave multiple scattering in the last decade. The coherent approach can have important differences from the incoherent approach in vegetation because the scatterers of vegetation are not uniformly randomly distributed. For example, leaves are usually in clusters and the positions of branches follow some architectural patterns. This means that scattering fields from different branches or

leaves can have definite relative phases. Thus, a coherent approach is sometimes necessary, particularly for long wavelengths such as L-band.

Coherent superposition

A first-order approximation is also known as the coherent superposition approximation, or coherent addition approximation (CAA), in which the electric fields from different scattering objects are summed together coherently (Zhang et al., 1996). For N scattering objects, the electric field of the scattered wave is expressed as

$$\bar{E} = \bar{E}_1 + \bar{E}_2 + \cdots + \bar{E}_N \quad (12)$$

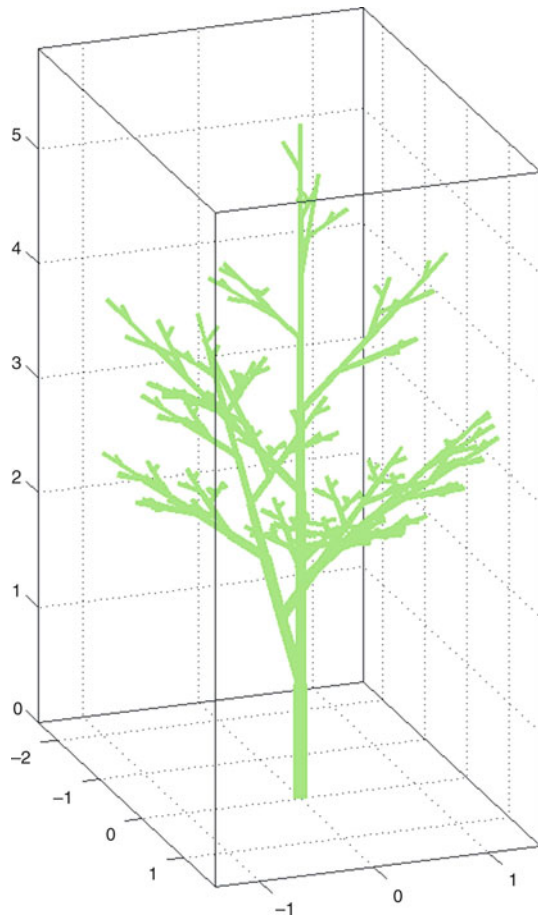
where $\bar{E}_i$ is the electric field of the scattered wave from the i^{th} object. The scattered intensity is proportional to $|\bar{E}|^2$. Thus, the relative phases of the scattered waves from different objects are included in the intensity. In actual computational implementations, it is necessary to define the positions of the branches and leaves. In vegetation scattering simulations, a common approach is to use computer growth mechanisms such as the Lindenmayer systems (L-systems) of tree growth (Chen et al., 1995; Tsang et al., 2001). An example of treelike structure generated by using the L-systems is shown in Figure 14.

In Figure 15, we show the backscattering coefficients of σ_{vv} and σ_{hh} , respectively, for the active remote sensing case based on the coherent superposition approach. The frequency is at 1.5 GHz. The trees are generated using the L-systems and the maximum tree height is about 1.64 m. The fractional volume occupied by the branches in the forest of trees is 0.12 %. The dielectric constant of branches is $11 + i4$ and that of the ground is $16 + i4$. The differences of the co-polarizations between the coherent addition approximation and the independent scattering approximation can be larger than 3 dB, especially for the horizontal polarization. One can observe some clustering effects since the distance between neighboring branches can be smaller than one wavelength and the scatterers' positions are correlated.

Full-wave multiple scattering

Scattering of electromagnetic waves by a layer of vertical cylinders overlying a ground surface has useful applications in both active and passive remote sensing of vegetation canopies (Tsang et al., 1995). For example, in forests and cornfields, the dominant scattering structures of plants are the trunks and cornstalks. The Foldy-Lax multiple scattering equations have been applied with the use of half-space Green's function to take into account the ground reflection effects. The half-space Green's function is expressed in terms of vector cylindrical waves. The Foldy-Lax scattering equations are solved to the second order to account for the clustering effects of cylinders in close proximity with each other.

In Figure 16, we show the bistatic scattering results for vertical cylinders in the clustered distribution case, where

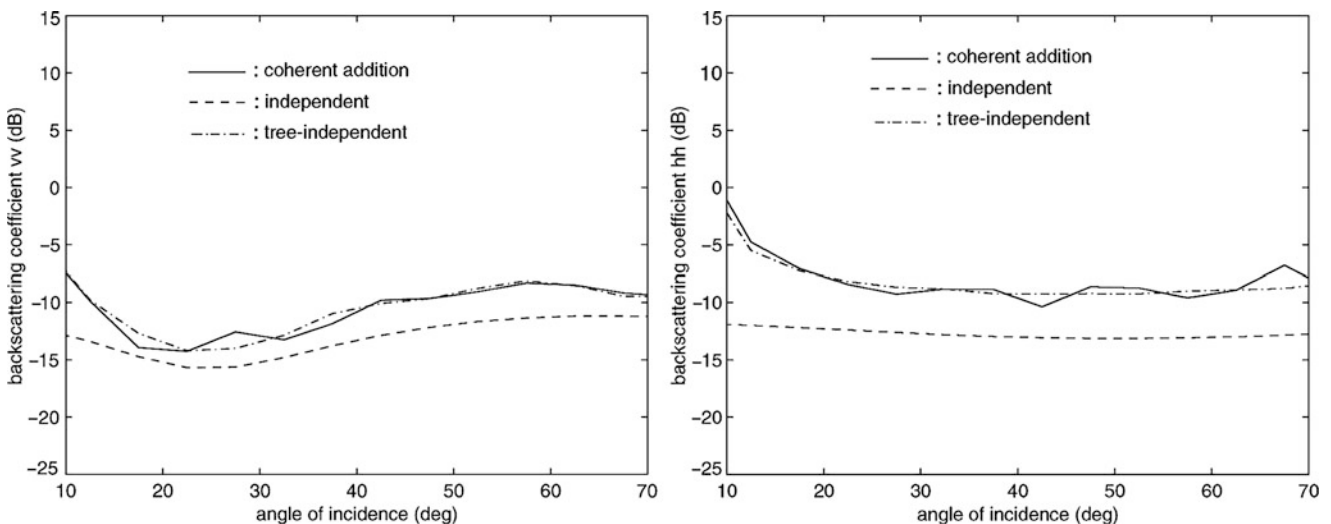


Radiation, Volume Scattering, Figure 14 Example of tree-like structure generated using the L-systems.

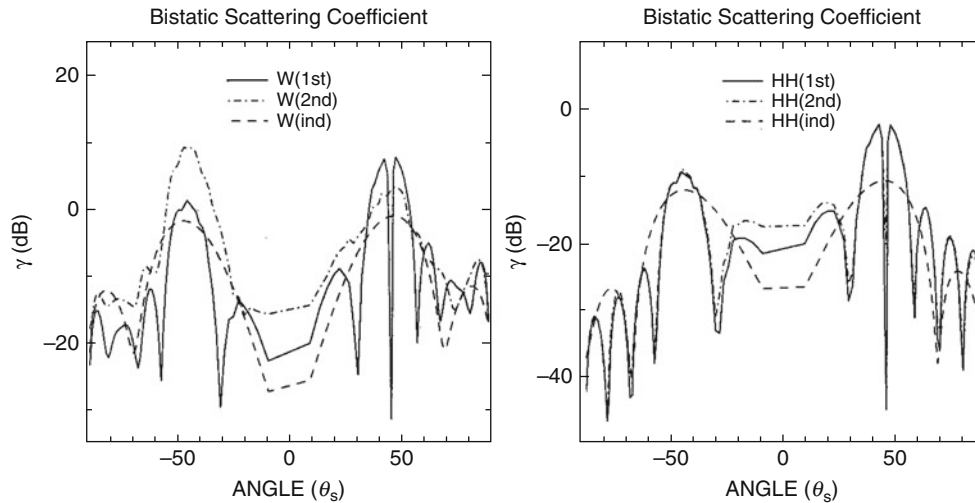
the second-order solution becomes important. The total fractional volume of cylinders is 0.1 %. We note that the second-order scattering is much larger than the first-order scattering. The second-order scattering of vertical polarization is more important than that of horizontal polarization. That is because vertical cylinders scatter vertically polarized waves more than horizontally polarized waves and create more coherent wave interaction among the cylinders in a cluster.

An important application of the vertical cylinder scattering model is in the remote sensing of paddy rice. The rice field is featured with a flooded ground surface during a large portion of its growing period and a nearly vertical rice plant structure. The rice crop scattering model consists of four major backscattering mechanisms: (a) direct scattering from the scatterer, (b) single scattering from the scatterer followed by reflection off the ground surface, (c) ground surface reflection followed by single scattering from the scatterer, and (d) reflection by ground surface followed by single scattering from the scatterer and further reflection off the ground surface, which are shown in Figure 12.

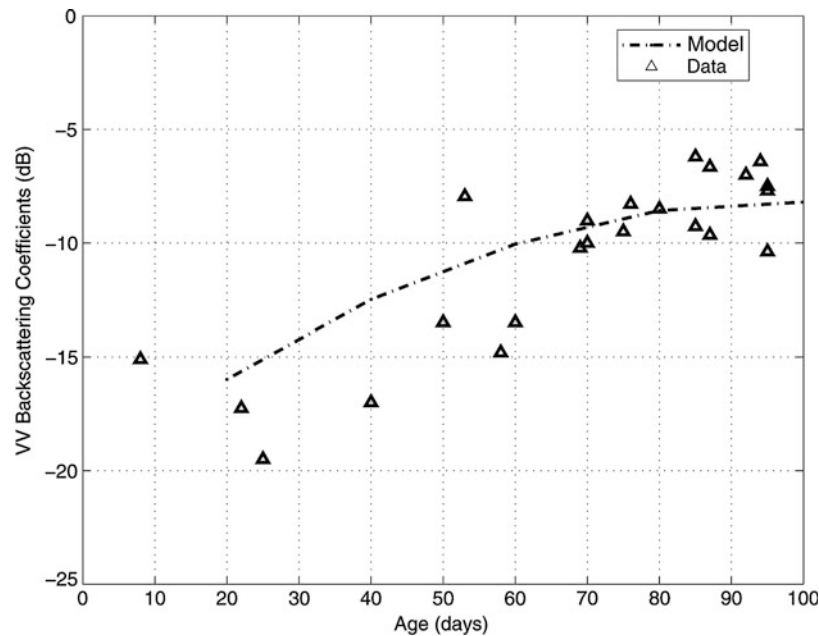
The rice canopy scattering model has been applied to simulate the backscattering signatures at different rice growth stages and compared with ERS-1 data. In Figure 15, the model backscattering results at different growth stages are compared with the ERS-1 vertical polarization data. The frequency is 5.3 GHz and the incidence angle is 23° . The comparison shows good agreement between model and measurements. The increasing trend of the temporal radar response is well captured by the model. Since the bottom of rice plants is immersed in water, the backscattering returns are dominated by the volume-surface interactions (Figure 17).



Radiation, Volume Scattering, Figure 15 Backscattering coefficients σ_{vv} (left) and σ_{hh} (right) of coherent addition, tree-independent scattering, and independent scattering approximations.



Radiation, Volume Scattering, Figure 16 Comparison of bistatic scattering coefficients (in dB) of VV (left) and HH (right). Results of the first-order, the second-order, and the independent scattering are shown.



Radiation, Volume Scattering, Figure 17 Comparison of measured and model backscattering coefficients at C-band.

Summary

This entry describes the volume scattering which occurs inside a medium containing scatterers with discrete permittivity. The electromagnetic wave will be absorbed and scattered in various directions when it propagates through the medium. Because of the complicated nature of geophysical media and the need of experimental data interpretation and geophysical parameters retrieval for remote sensing, this entry presents a physically based vector radiative transfer model and the full-wave approach for volume scattering in dense media and

vegetation canopy. This entry also provides some validations between the forward scattering model simulations and microwave remote sensing measurements in snow and vegetation and illustrates the usefulness of these volume scattering models. Through the advent of computer technology and fast computational algorithm development, and better understanding of scattering physics in geophysical media, continuous progress has been made in improving electromagnetic wave propagation and scattering models to assist in data interpretation and geophysical inversion.

Bibliography

- Chen, Z., Tsang, L., and Zhang, G., 1995. Scattering of electromagnetic waves by vegetation based on the wave approach and Stochastic Lindenmayer system. *Microwave and Optical Technology Letters*, **8**(1), 30–33.
- Chen, C. T., Tsang, L., Guo, J., Chang, A. T. C., and Ding, K.-H., 2003. Frequency dependence of scattering and extinction of dense media based on three-dimensional simulations of Maxwell's equations with applications to snow. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 1844–1852.
- Dobson, C., Stiles, H., Brunfeldt, D., Metzler, T., and McMeeking, S., 1977. *Data Documentation: 1975 MAS 8–18 Vegetation Experiments*. Report of RSL TR 264–15, Remote Sensing Laboratory, University of Kansas Center for Research, Lawrence.
- Hornbuckle, B. K., England, A. W., De Roo, R. D., Fischman, M. A., and Boprie, D. L., 2003. Vegetation canopy anisotropy at 1.4 GHz. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(10), 2211–2223.
- Le Vine, D. M., and Karam, M. A., 1996. Dependence of attenuation in a vegetation canopy on frequency and plant water content. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(5), 1090–1096.
- Liang, D., Xu, X., Tsang, L., Andreadis, K. M., and Josberger, E. G., 2008. Multi-layer effects in passive microwave remote sensing of dry snow using dense media radiative transfer theory (DMRT) based on quasicrystalline. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(11), 3663–3671.
- McDonald, K. C., Dobson, M. C., and Ulaby, F. T., 1990. Using mimics to model L-band multiangle and multitemporal backscatter from a walnut orchard. *IEEE Transactions on Antennas and Propagation*, **28**(4), 477–491.
- Owe, M., de Jeu, R., and Walker, J., 2001. A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(8), 1643–1654.
- Parde, M., Wigneron, J.-P., Waldteufel, P., Kerr, Y. H., Chanzy, A., Sobaerg, S. S., and Skou, N., 2004. N-parameter retrievals from l-band microwave observations acquired over a variety of crop fields. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(6), 1168–1178.
- Tsang, L., and Kong, J. A., 2001. *Scattering of Electromagnetic Waves*. Hoboken: Wiley-Interscience. Advanced Topics, Vol. 3.
- Tsang, L., and Li, Q., 2004. Wave scattering with UV multilevel partitioning method for volume scattering by discrete scatterers. *Microwave and Optical Technology Letters*, **41**(5), 354–361.
- Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley-Interscience.
- Tsang, L., Chan, C. H., Kong, J. A., and Joseph, J., 1992. Polarimetric signatures of a canopy of dielectric cylinders based on first- and second-order vector radiative transfer theory. *Journal of Electromagnetic Waves and Applications*, **6**(1), 19–51.
- Tsang, L., Ding, K.-H., Zhang, G., Hsu, C. C., and Kong, J. A., 1995. Backscattering enhancement and clustering effects of randomly distributed dielectric cylinders overlying a dielectric half space based on Monte-Carlo simulations. *IEEE Transaction on Antennas and Propagation*, **43**(5), 488–499.
- Tsang, L., Kong, J. A., and Ding, K.-H., 2000. *Scattering of Electromagnetic Waves*. Hoboken: Wiley-Interscience. Theory and Applications, Vol. 1.
- Tsang, L., Kong, J. A., Ding, K.-H., and Ao, C. O., 2001. *Scattering of Electromagnetic Waves*. Hoboken: Wiley-Interscience. Numerical Simulations, Vol. 2.
- Tsang, L., Pan, J., Liang, D., Li, Z. X., Cline, D., and Tan, Y. H., 2007. Modeling active microwave remote sensing of snow using dense media radiative transfer (DMRT) theory with multiple scattering effects. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(4), 990–1004.
- Tse, K. K., Tsang, L., Chan, C. H., Ding, K.-H., and Leung, K. W., 2007. Multiple scattering of waves by dense random distribution of sticky particles for applications in microwave scattering by terrestrial snow. *Radio Science*, **42**, RS5001.
- Ulaby, F. T., and Wilson, E. A., 1985. Microwave attenuation properties of vegetation canopies. *IEEE Transactions on Geoscience and Remote Sensing*, **23**(5), 746–753.
- Ulaby, F. T., Sarabandi, K., McDonald, K., Whitt, M., and Dobson, M. C., 1990. Michigan microwave canopy scattering model. *International Journal of Remote Sensing*, **11**(7), 1223–1253.
- Van de Griend, A. A., and Wigneron, J.-P., 2004. The factor b as a function of frequency and canopy type at H-polarization. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(4), 786–794.
- Van de Griend, A. A., Owe, M., De Ruiter, J., and Gouweleeuw, B. T., 1996. Measurement and behavior of dual-polarization vegetation optical depth and single scattering albedo at 1.4- and 5-GHz microwave frequencies. *IEEE Transactions on Geoscience and Remote Sensing*, **34**(4), 957–965.
- Wigneron, J.-P., Parde, M., Waldteufel, P., Chanzy, A., Kerr, Y., Schmidl, S., and Skou, N., 2004. Characterizing the dependence of vegetation model parameters on crop structure, incidence angle, and polarization at L-band. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(2), 416–425.
- Yueh, S. H., Kong, J. A., Jao, J. K., Shin, R. T., and Le Toan, T., 1992. Branching model for vegetation. *IEEE Transactions on Geoscience and Remote Sensing*, **30**(2), 390–402.
- Zhang, G., Tsang, L., and Chen, Z., 1996. Collective scattering effects of trees generated by Stochastic Lindenmayer systems. *Microwave and Optical Technology Letters*, **11**(2), 107–111.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)
[Fields and Radiation](#)
[Geophysical Retrieval, Overview](#)
[Land Surface Emissivity](#)
[Media, Electromagnetic Characteristics](#)
[Microwave Surface Scattering and Emission](#)
[Radars](#)
[Radiation, Electromagnetic](#)
[Radiation, Multiple Scattering](#)
[Radiative Transfer, Solution Techniques](#)
[Radiative Transfer, Theory](#)
[Remote Sensing, Physics and Techniques](#)
[Terrestrial Snow](#)

RADIATIVE TRANSFER, SOLUTION TECHNIQUES

Rodolfo Guzzi

Agenzia Spaziale Italiana ASI, Roma, Italy

Definitions

- $A(\zeta)$ Surface albedo
- A^m Sobolev's scattering functions
- B Source function
- B_1^* Source function from above
- B_*^m Source function from below
- $B^m, I^m, I_{sur}^m, p^m, y^m$ Fourier expansions

- $C_{\pm jp}$ Coefficients for Discrete Ordinate Method
- $f = Kf + \Psi$ Neumann series
- f^+ Adjoint RTE
- g Heney-Greenstein parameter
- $g(t)$ Green's function
- k_p Transport transition density
- I Intensity
- I_{sur} Intensity function reflected by the underlying surface
- I^1 Single bottom light intensity
- I_1^{*m} Sun intensity due to primary scattering from above
- I_{*1}^m Sun illumination from below the atmosphere
- $I_{sur,1}^{*m}$ Direct solar photons reflected by underlying surface
- L Current discretization
- L_x Mathematics operator
- M High signal frequency
- $P(\mathbf{r}, \omega, \omega')$ Phase function used for Monte Carlo
- $P_i(\cos\gamma)$ Legendre's polynomials
- P_i^m Associate Legendre's polynomials
- pdf Probability density function
- $r(\eta)$ Fresnel coefficient of mirror reflection
- r_{sur} Surface phase function
- r_n Random number generator
- $R_{i,m}^m$ Sobolev's polynomials
- S Incident monochromatic solar radiation
- γ Arbitrary probabilistic function
- $\beta(\mathbf{r})$ Extinction coefficient
- χ Phase function or scattering function
- χ_{sur} Surface phase function
- δ_{0m} Delta Kronecker
- δ Delta Dirac
- η' Cosine of scattered beam
- η'' Cosine of incident beam
- $\in$ Level of error
- $\tilde{\in}$ Interpolation accuracy
- $\in^*$ Second level of error
- γ Scattering angle
- λ Wavelength
- Λ Single scattering albedo
- $d\omega$ Increment of solid angle
- ω_j Gaussian weight
- Ω Solid angle
- $\Psi_{solar}, \Psi_{thermal}$ Source functions used in Monte Carlo
- τ Optical depth
- τ_0 Optical thickness
- ϕ Azimuth angle
- ϕ_0 Sun azimuth angle
- ϕ', ϕ'' Azimuth angles
- ζ Cosine of Sun incident angle

Introduction

From the pioneering books by Chandrasekhar (1960) and Sobolev (1975) to today, many different books on light propagation in the atmosphere have been written. All these works investigate different methods and algorithms to obtain solutions suitable for use as application models. We can divide the methods into three categories. The first

is based on analytical solutions and includes, in addition to Chandrasekhar and Sobolev, and Yanovitsky (1997), the work done by Van de Hulst (1980) on double adding and on its implemented code, done by DeHaan et al. (1987), which also contains polarization computation, and the method developed by Kondratiev et al. (1992).

The second category includes numerical solution methods, starting from the two books by Liou (1980, 1992) up to the most recent by Thomas and Stamnes (1999), Bohren and Clothiaux (2006), a revised version of the book by Petty (2006), Goody and Young (1989), and Lenoble (1993).

The third category consists of statistical modeling or Monte Carlo methods, for example, the books by Marchuk et al. (1980), Preisendorfer (1965), and the most recent for 3D in cloudy atmosphere by Marshak and Davis (2005). A valuable work was also done by the I3RC group (<http://code.google.com/p/i3rc-monte-carlo-model/>).

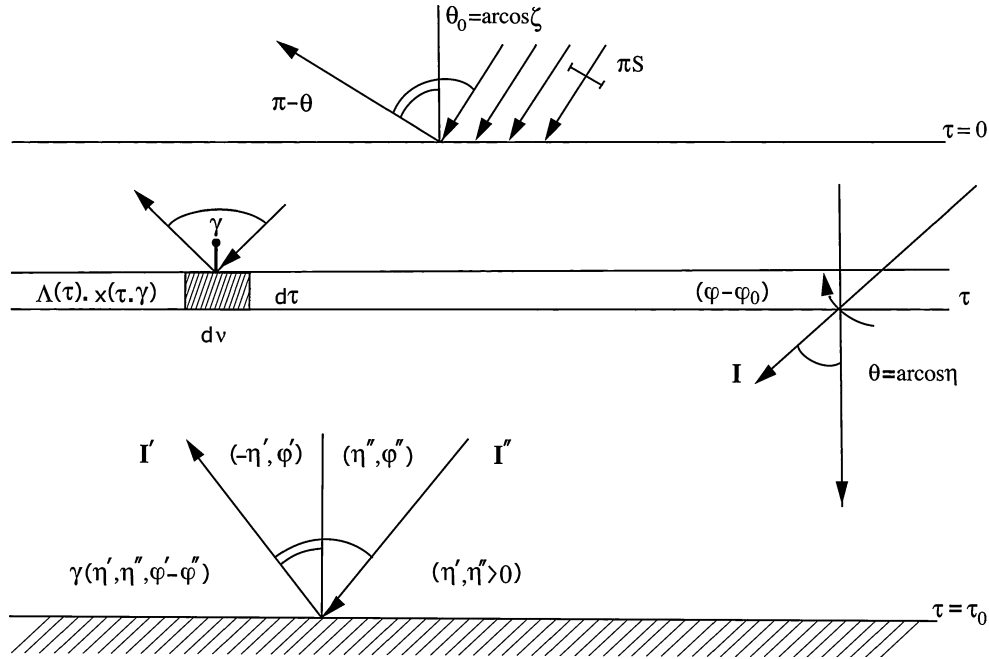
In 1975 and 1980, the International Association of Meteorology and Atmospheric Physics (IAMAP) Radiation Commission released two documents (1975; 1980) describing the different solutions of the radiative transfer equations in order to investigate the accuracy of the different methods proposed and to find the most economical procedures for various future applications. In 1985, Lenoble (1985) wrote a more complete book in which several solution methods were described in more detail and a numerical comparison of the methods was shown.

At present, it is not easy to define which method is more precise than another. New methods should not only overcome the present limitations, mainly linked to the geometry and dimensions of the system, but also should be able to capture the real interactions between the scattering and absorption by molecule and aerosol. Models of radiation are becoming increasingly sophisticated due to hyperspectral sensors mounted on board satellites which have high spatial resolution.

Radiative transfer: initial boundary value problem

Let us consider a photon beam traveling into a vertically nonuniform atmosphere containing aerosols and clouds. Bounded from a non-orthotropic reflecting bottom, it produces complex phenomena of multiple scattering and emissions requiring the simultaneous solution of two equations applied to an atmosphere-surface system. One of these is the radiative transfer, which analyzes the intensity change of radiation, while the other expresses the emission in terms of radiative equilibrium. The light theory of scattering in planetary atmosphere can be formulated with the aid of these equations. The solution of these equations is the topic of this entry.

Now let us consider a plane-parallel planetary atmosphere of a finite optical thickness τ_0 overlying an arbitrary reflecting surface. The primary scattering and absorption process somewhere in the atmosphere at any optical depth τ and each fixed wavelength λ is described by the albedo of single scattering $\Lambda(\tau)$ and by the



Radiative Transfer, Solution Techniques, Figure 1 Definition of geometrical variables and optical quantities employed to describe multiple light scattering.

phase function $\chi(\tau, \eta', \eta'', \phi' - \phi'')$. If we define, with respect to a downward normal, the cosine of the scattered beam with variable angles η' , and that of the light beams incident with η'' , and azimuthal angles with the variables ϕ' and ϕ'' measured clockwise looking from the bottom to the top of the atmosphere, then the proper scattering angles are denoted by γ , which is a function of the above-defined angular variables

$$\cos \gamma = \eta' \eta'' + \sqrt{(1 - \eta'^2)(1 - \eta''^2)} \cos(\phi' - \phi'')$$

Since it is well known that the real primary optical characteristics of planet's atmospheres and natural underlying surfaces are strongly variable in space-time scales, we shall consider vertical nonuniformities and bottom reflectance laws arbitrarily. Our approach to the problem, at first, is to neglect horizontal nonuniformities of optical atmospheric parameters, besides refraction and polarization effects.

Let us suppose now that the top of this planetary atmosphere is uniformly illuminated by parallel solar rays in the direction (ζ, ϕ_0) and that the incident monochromatic net flux of solar radiation per a unit area normal to the direction of solar ray propagation is equal to πS . The intensity of radiation in the current direction (η, ϕ) at any atmospheric optical depth τ will be denoted by $I(\tau; \eta, \phi - \phi_0, \tau_0)$, whereas all current angles $\arccos \zeta$ and $\arccos \eta$ are measured, respectively, for the inward normal of a slab, as shown in Figure 1. The wavelength index λ , depending on all abovementioned optical quantities, will be omitted for simplicity.

The law of bidirectional light reflectance from underlying surfaces is described by the arbitrary probabilistic function $y(\eta', \eta'', \phi' - \phi'')$, which transforms the intensity I'' of incident radiation from the direction η'', ϕ'' into the intensity I' of reflected radiation in the direction $-\eta', \phi'$ (Preisendorfer, 1965):

$$I' \eta' = \frac{1}{2\pi} y(\eta', \eta'', \phi' - \phi'') I'' \eta'' \quad (1)$$

The planet surface is illuminated by both unscattered but attenuated solar radiation and by diffuse radiation from atmosphere.

The probabilistic function introduced in the classical radiative transfer theory by V. Sobolev (1975) can be also used for the surface phase function $r_{sur}(\tau, \eta', \eta'', \phi' - \phi'')$ according to the following relation:

$$2\eta' r_{sur}(\eta', \eta'', \phi' - \phi'') = y(\eta', \eta'', \phi' - \phi'') \quad (2)$$

Due to the probabilistic character of function $y(\eta', \eta'', \phi' - \phi'')$, the normalized condition has to be taken into account:

$$\frac{1}{2\pi} \int_{\Omega} y(\eta', \eta'', \phi' - \phi'') = 1 \quad (3)$$

We need to emphasize that, unlike atmospheric and surface phase functions, respectively, defined as $\chi(\eta', \eta'', \phi' - \phi'')$ and $r_{sur}(\eta', \eta'', \phi' - \phi'')$, the probabilistic function $y(\eta', \eta'', \phi' - \phi'')$ does not possess the angular symmetry properties expressed by the principle of

optical reciprocity. Namely, we have the following disequality:

$$y(\eta', \eta'', \varphi' - \varphi'') \neq y(\eta'', \eta', \varphi' - \varphi'') \quad (4)$$

With these statements in mind, let us formulate the initial boundary value problem of the radiative transfer theory needed for the determination of unknown intensities $I(\tau, \eta, \zeta, \varphi - \varphi_0, \tau_0)$.

According to well-known classical considerations (see, e.g., Sobolev, 1975), the intensity is a solution of the following initial boundary value problem:

$$\begin{aligned} \eta \frac{dI(\tau, \eta, \zeta, \varphi - \varphi_0, \tau_0)}{d\tau} = & -I(\tau, \eta, \zeta, \varphi - \varphi_0, \tau_0) \\ & + B(\tau, \eta, \zeta, \varphi - \varphi_0, \tau_0) \\ & \tau \in [0, \tau_0], \eta \in [-1, 1], \\ & \zeta \in [0, 1] \end{aligned} \quad (5)$$

where the source function $B(\tau, \eta, \zeta, \varphi - \varphi_0, \tau_0)$ can be represented in integral form:

$$\begin{aligned} B(\tau, \eta, \zeta, \varphi - \varphi_0, \tau_0) = & \frac{\Lambda(\tau)}{4\pi} \int_0^{2\pi} d\varphi' \int_{-1}^1 \chi(\tau, \eta, \eta', \varphi - \varphi') \\ & I(\tau, \eta, \varphi' - \varphi_0, \tau_0) d\eta' \\ & + \frac{S}{4} \Lambda(\tau) \chi(\tau, \eta, \zeta, \varphi - \varphi_0) \exp\left(-\frac{\tau}{\zeta}\right) \\ & \tau \in [0, \tau_0], \eta \in [-1, 1], \zeta \in [0, 1] \end{aligned} \quad (6)$$

The boundary conditions for the given boundary value problem are:

$$\begin{aligned} I(0, \eta, \zeta, \varphi - \varphi_0, \tau_0) &= 0 \quad \eta \in [0, 1], \zeta \in [0, 1] \\ I(\tau_0, -\eta, \zeta, \varphi - \varphi_0, \tau_0) &= I_{sur}(\eta, \zeta, \varphi - \varphi_0, \tau_0) \end{aligned} \quad (7)$$

where I_{sur} is the intensity of radiation, multiply reflected by the underlying surface at $\tau = \tau_0$. Taking into account Eq. 1, we get the following integral expression for $I_{sur}(\eta, \zeta, \varphi - \varphi_0, \tau_0)$, as also it is given by Sobolev (1975):

$$\begin{aligned} \eta I_{sur}(\eta, \zeta, \varphi - \varphi_0, \tau_0) = & \frac{1}{2\pi} \int_0^{2\pi} d\varphi' \int_0^1 y(\eta, \eta', \varphi - \varphi_0) \\ & I(\tau_0, \eta', \zeta, \varphi' - \varphi_0, \tau_0) \eta' d\eta' \\ & + \frac{S}{2} y(\eta, \zeta, \varphi - \varphi_0) \zeta \exp\left(-\frac{\tau_0}{\zeta}\right) \\ & \eta \in [0, 1], \zeta \in [0, 1] \end{aligned} \quad (8)$$

The formal solutions of the radiative transfer Eq. 5, taking into account the expression for the source function

Eq. 6, can be written directly in integral form, satisfying the boundary conditions Eq. 7:

$$\begin{aligned} I(\tau, \eta, \varphi - \varphi_0, \tau_0) = & \int_0^\tau B(\tau', \eta, \zeta, \varphi - \varphi_0, \tau_0) \\ & \exp\left(-\frac{\tau - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \end{aligned} \quad (9)$$

$$\begin{aligned} I(\tau - \eta, \varphi - \varphi_0, \tau_0) = & \int_0^\tau B(\tau', -\eta, \zeta, \varphi - \varphi_0, \tau_0) \\ & \exp\left(-\frac{\tau' - \tau}{\eta}\right) \frac{d\tau'}{\eta} \\ & + I_{sur}(\eta, \zeta, \varphi - \varphi_0, \tau_0) \exp\left(-\frac{\tau_0 - \tau}{\eta}\right) \\ & \tau \in [0, \tau_0], \eta \in [0, 1], \zeta \in [0, 1] \end{aligned} \quad (10)$$

Substituting Eqs. 9 and 10 in Eq. 6 and taking into account different mathematical aspects of upwelling and downwelling radiation, we can obtain the well-known system of exact integral equations for the source function B.

On the contrary, by substituting Eq. 6 in Eqs. 9 and 10, the appropriate exact integral equations for upwelling and downwelling intensities can be obtained.

Fourier series expansions

Since in the theory of light scattering the phase function is frequently expanded in Legendre polynomials, we can separate the azimuthal dependence on I and B , expanding these quantities in a Fourier series. Then the system of Eqs. 5 and 6 can be separated into distinct pairs of equation determining the coefficient of this expansion. Then we may use azimuth's harmonics of Fourier series expansions for the following optical quantities:

$$\begin{aligned} \chi(\tau, \gamma) = & \sum_{i=0}^{M_1} x_i(\tau) P_i(\cos \gamma) = p^0(\tau, \eta, \zeta) \\ & + 2 \sum_{m=1}^{M_1} p^m(\tau, \eta, \zeta) \cos m(\varphi - \varphi_0) \end{aligned} \quad (11)$$

$$\begin{aligned} y(\eta', \eta'', \varphi' - \varphi'') = & y^0(\eta', \eta'') \\ & + 2 \sum_{m=1}^{M_2} y^m(\tau, \eta', \eta'') \cos m(\varphi - \varphi'), \end{aligned} \quad (12)$$

$$\begin{aligned} I(\tau, \eta, \zeta, \varphi - \varphi_0, \tau_0) = & I^0(\tau, \eta, \zeta, \tau_0) \\ & + 2 \sum_{m=1}^{[M_1, M_2]} I^m(\tau, \eta, \zeta, \tau_0) \cos m(\varphi - \varphi_0) \end{aligned} \quad (13)$$

$$I(\tau, -\eta, \zeta, \varphi - \varphi_0, \tau_0) = I^0(\tau, -\eta, \zeta, \tau_0) + 2 \sum_{m=1}^{[M_1, M_2]} I^m(\tau, -\eta, \zeta, \tau_0) \cos m(\varphi - \varphi_0) \quad (14)$$

with $m = 0, 1, 2, \dots, M_1$ and $m = 0, 1, 2, \dots, M_2$ and where $P_i(\cos \gamma)$ are Legendre polynomials. From here forward, only azimuth's number $M = \max[M_1, M_2]$ will be used for the Fourier analysis of the optical atmosphere-surface quantities. The functions $p^m(\tau, \eta, \zeta)$ are calculated on the basis of the well-known series expansion or with the help of traditional Fourier representation:

$$p^m(\tau, \eta, \zeta) = \sum_{i=m}^{M_i} x_i(\tau) \frac{(m-i)!}{(m+i)!} P_i^m(\eta) P_i^m(\zeta) \quad (15)$$

$$p^m(\tau, \eta, \zeta) = \frac{1}{2\pi} \int_0^{2\pi} \chi(\tau, \eta, \zeta, \varphi - \varphi_0) \cos m \times (\varphi - \varphi_0) d(\varphi - \varphi_0) \quad (16)$$

The functions $P_i^m(\eta)$ are the associated Legendre's polynomials, and coefficients x_i are calculated according to:

$$x_i(\tau) = \frac{2i+1}{2} \int_0^{2\pi} \chi(\tau, \gamma) P_i(\cos \gamma) \sin \gamma d\gamma |x_i(\tau)| \leq 2i+1; i=0, 1..M \quad (17)$$

Of course, the atmospheric phase functions $\chi(\tau, \gamma)$ will be normalized in accordance with the usual conditions:

$$\frac{1}{2} \int_0^\pi \chi(\tau, \gamma) \sin \gamma d\gamma = 1 \quad (18)$$

Taking into account Eqs. 11, 12, 13, and 14, we obtain the basic boundary value problem for the determination of azimuth's harmonics radiation fields:

$$\eta \frac{dI^m(\tau, \eta, \zeta, \tau_0)}{d\tau} = -I^m(\tau, \eta, \zeta, \tau_0) + B^m(\tau, \eta, \zeta, \tau_0) \quad \tau \in [0, \tau_0], \eta \in [-1, 1], \zeta \in [0, 1] \quad (19)$$

$$B^m(\tau, \eta, \zeta, \tau_0) = \frac{\Lambda(\tau)}{4} \int_{-1}^1 p^m(\tau, \eta, \eta', \tau_0) I^m(\tau, \eta, \zeta, \tau_0) d\eta' + \frac{S}{4} \Lambda(\tau) p^m(\tau, \eta, \zeta) \exp\left(-\frac{\tau}{\eta}\right) \quad \tau \in [0, \tau_0], \eta \in [-1, 1], \zeta \in [0, 1] \quad (20)$$

with the boundary conditions

$$I^m(0, \eta, \zeta, \tau_0) = 0 \quad \eta \in [0, 1], \zeta \in [0, 1] \quad (21)$$

$$I^m(\tau_0, -\eta, \zeta, \tau_0) = I_{sur}^m(\eta, \zeta, \tau_0)$$

The azimuth's harmonics of the surface brightness I_{sur} can be expanded as

$$\eta I_{sur}^m(\eta, \zeta, \tau_0) = \frac{S\zeta}{2} y^m(\eta, \zeta) \exp\left(-\frac{\tau_0}{\zeta}\right) + \int_0^1 y^m(\eta, \eta') I^m(\tau_0, \eta', \zeta, \tau_0) \eta' d\eta' \quad (22)$$

where y^m has been obtained by expanding it in the same way of P^m .

Then, the general integral forms with the previous boundary conditions are

$$I^m(\tau, \eta, \zeta, \tau_0) = \int_0^\tau B^m(\tau', \eta, \zeta, \tau_0) \exp\left(-\frac{\tau - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \quad (23)$$

$$I^m(\tau, -\eta, \zeta, \tau_0) = \int_0^\tau B^m(\tau', -\eta, \zeta, \tau_0) \exp\left(-\frac{\tau' - \tau}{\eta}\right) \frac{d\tau'}{\eta} + I_{sur}^m(\eta, \zeta, \tau_0) \exp\left(-\frac{\tau_0 - \tau}{\eta}\right) \quad (24)$$

According to the classical approach, substituting Eqs. 23 and 24 into Eq. 20, we can obtain the exact integral equations for the azimuth's harmonics of the source function $B^m(\tau, \eta, \zeta, \tau_0)$ for upward and downward directions:

$$B^m(\tau, \eta, \zeta, \tau_0) = \frac{\Lambda(\tau)}{2} \left[\int_0^1 p^m(\tau, \eta, \eta') I^m(\tau, \eta', \zeta, \tau_0) d\eta' + \int_0^1 p^m(\tau, -\eta, \eta') I^m(\tau, -\eta, \zeta, \tau_0) d\eta' \right] + \frac{S}{4} \Lambda(\tau) p^m(\tau, \eta, \zeta) \exp\left(\frac{-\tau}{\zeta}\right) \quad (25)$$

$$B^m(\tau, -\eta, \zeta, \tau_0) = \frac{\Lambda(\tau)}{2} \left[\int_0^1 p^m(\tau, -\eta, \eta') I^m(\tau, \eta', \zeta, \tau_0) d\eta' + \int_0^1 p^m(\tau, \eta, \eta') I^m(\tau, -\eta, \zeta, \tau_0) d\eta' \right] + \frac{S}{4} \Lambda(\tau) p^m(\tau, -\eta, \zeta) \exp\left(\frac{-\tau}{\zeta}\right) \quad (26)$$

From an applied point of view, it is better to obtain the proper exact integral equation from azimuth's harmonics $I^m(\tau, \pm\eta, \zeta, \tau_0)$. Substituting Eqs. 25 and 26 into Eqs. 23 and 24, we obtain the integral forms for the radiative transfer equations:

$$I^m(\tau, \eta, \zeta, \tau_0) = \int_0^{\tau_0} \frac{\Lambda(\tau')}{2} \left\{ \left[\int_0^1 p^m(\tau, \eta, \eta') I^m(\tau, \eta', \zeta, \tau_0) d\eta' \right. \right. \\ \left. \left. + \int_0^1 p^m(\tau, -\eta, \eta') I^m(\tau, -\eta, \zeta, \tau_0) d\eta' \right] \right. \\ \left. + \frac{S}{2} p^m(\tau, \eta, \zeta) \exp\left(-\frac{\tau'}{\zeta}\right) \right\} \exp\left(-\frac{\tau - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \quad (27)$$

and

$$I^m(\tau, -\eta, \zeta, \tau_0) = \int_0^{\tau_0} \frac{\Lambda(\tau')}{2} \left\{ \left[\int_0^1 p^m(\tau, -\eta, \eta') I^m(\tau, \eta', \zeta, \tau_0) d\eta' \right. \right. \\ \left. \left. + \int_0^1 p^m(\tau, -\eta, \eta') I^m(\tau, -\eta, \zeta, \tau_0) d\eta' \right] \right. \\ \left. + \frac{S}{2} p^m(\tau, -\eta, \zeta) \exp\left(-\frac{\tau'}{\zeta}\right) \right\} \exp\left(-\frac{\tau - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \\ + I_{sur}^m(\eta, \zeta, \tau_0) \exp\left(-\frac{\tau_0 - \tau}{\eta}\right) \quad (28)$$

These are the integral equations useful for the calibration. Unfortunately, classical analytical and semi-analytical approaches cannot provide a complete numerical analysis of radiation fields existing in the environment systems. Nevertheless, with the help of precise analytical and highly accurate semi-analytical methods and algorithms, we are able to calibrate the actual spatial-angular distributions of radiation fields in order to estimate the precision of the input environment optical models. In fact, the application of precise calibrating models of radiation fields is necessary in order to estimate the accuracy of different approximate algorithms, used in atmospheric correction and in the thematic interpretation of modern multispectral and hyperspectral satellite information.

It should be noted that functions $p^m(\tau, \eta, \zeta)$ can also be calculated without any information about coefficients x_i . In fact, with the help of known scattering functions $A^m(\tau, \eta, \zeta)$,

$$p^m(\tau, \eta, \zeta) = A^m(\tau, \eta, \zeta) \quad \text{for } \Lambda = 0 \quad (29)$$

$A^m(\tau, \eta, \zeta)$ can be calculated on the basis of the expansion series given by Sobolev (1975)

$$A^m(\tau, \eta, \zeta) = P_m^m(\eta) \sum_{i=m}^M x_i(\tau) \frac{(m-i)!}{(m+i)!} P_i^m(\zeta) R_{i,m}^m(\eta, \zeta) \quad (30)$$

where functions $R_{i,m}^m$ are Sobolev's polynomials, which can be calculated on the basis of recursive formulas:

$$(i-m+1)R_{i+1,m}^m(\eta) + (i+m)R_{i-1,m}^m(\eta) \\ = (2i+1\Lambda x_i)R_{i,m}^m(\eta) d\eta \quad (31)$$

where

$$R_{i,m}^m(\eta) = 0 \quad m > i \quad (32)$$

$$R_{i,m}^m(\eta) = 1 \quad m > i \quad (33)$$

It should be noted that in case of large $m \gg 1$ and for $\eta \approx 1$, $\eta \rightarrow 1$, $\zeta \rightarrow 1$, $A^m(\tau; \eta, \zeta)$ losses of accuracy due to significant oscillations of small value used for calculating the functions.

Successive approximation methods and the iterative scheme

The system of exact integral Eqs. 27 and 28 can be solved numerically on the basis of successive approximations method according to an iterative scheme. We have the downwelling radiation as

$$I_{n+1}^m(\tau, \eta, \zeta, \tau_0) = \int_0^{\tau_0} \frac{\Lambda(\tau')}{2} \left\{ \left[\int_0^1 p^m(\tau', \eta, \eta') I_n^m(\tau', \eta', \zeta, \tau_0) d\eta' \right. \right. \\ \left. \left. + \int_0^1 p^m(\tau', -\eta, \eta') I_n^m(\tau', -\eta, \zeta, \tau_0) d\eta' \right] \right. \\ \left. + \frac{S}{2} p^m(\tau', \eta, \zeta) \exp\left(-\frac{\tau'}{\zeta}\right) \right\} \exp\left(-\frac{\tau - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \quad (34)$$

While for the upwelling intensity, the iterative scheme is

$$I_{n+1}^m(\tau, -\eta, \zeta, \tau_0) = \int_0^{\tau_0} \frac{\Lambda(\tau')}{2} \left\{ \left[\int_0^1 p^m(\tau', -\eta, \eta') I_n^m(\tau', \eta', \zeta, \tau_0) d\eta' \right. \right. \\ \left. \left. + \int_0^1 p^m(\tau', \eta, \eta') I_n^m(\tau', -\eta, \zeta, \tau_0) d\eta' \right] \right. \\ \left. + \frac{S}{2} p^m(\tau', -\eta, \zeta) \exp\left(-\frac{\tau'}{\zeta}\right) \right\} \exp\left(-\frac{\tau - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \\ + I_{sur}^m(\eta, \zeta, \tau_0) \exp\left(-\frac{\tau_0 - \tau}{\eta}\right) \quad (35)$$

The iterative brightness of underlying surface is

$$\eta I_{sur,n}^m(\eta, \zeta, \tau_0) = \frac{S\zeta}{2} y^m(\eta, \zeta) \exp\left(-\frac{\tau_0}{\zeta}\right) \quad (36)$$

$$+ \int_0^1 y^m(\eta, \eta') I_n^{*m}(\tau_0, \eta', \zeta) \eta' d\eta' \quad (37)$$

with $\tau \in [0, \tau_0]$ $\eta \in [0, 1]$ $\zeta \in [0, 1]$ and $m = 0, 1 \dots M$, $n = 1, 1, 2 \dots, K$.

The starting approximations for $I^m(\tau; \eta, \zeta, \tau_0)$ and $I^m(\tau, -\eta, \zeta, \tau_0)$ can be chosen equal, for example, according to the following scheme:

$$I_0^m(\tau, \eta, \zeta, \tau_0) = 0 \quad (38)$$

$$I_0^m(\tau, -\eta, \zeta, \tau_0) = 0 \quad (39)$$

In particular case, we can obtain the corresponding approximate expressions for the primary light scattering by the atmosphere and the single light reflection from the bottom as a result of the first iterative step. In order to get precise values of the primary atmospheric light reflection and the single light reflection coming from the bottom determined by the brightness $I_{sur,1}^m(\tau, \eta, \zeta, \tau_0)$ and $I_{sur,1}^m(\tau, -\eta, \zeta, \tau_0)$, we need to use this starting approximation:

$$I_0^m(\tau, \eta, \zeta, \tau_0) = 0 \quad (40)$$

$$I_0^m(\tau, -\eta, \zeta, \tau_0) = I_{sur,1}^m(\eta, \zeta, \tau_0) \exp\left(-\frac{\tau_0 - \tau}{\eta}\right) \quad (41)$$

where the bottom brightness $I_{sur,1}^m(\eta, \zeta, \tau_0)$ is determined by proper approximation for I_{sur}^m selected from the boundary condition Eq. 21, taking into account all processes of the primary atmospheric light scattering before and after single light reflection from the underlying surface. We emphasize that, especially from the point of view of calibration, the starting approximations must be chosen according to some arbitrary but exact or highly precise solutions of the radiative transfer equation. In particular, we can use, as the first starting step, the precise expressions that take into account exactly the primary atmospheric light scattering and single bottom light reflection:

$$I_0^m(\tau, \eta, \zeta, \tau_0) = I_1^m(\tau, \eta, \zeta, \tau_0) \quad (42)$$

$$I_0^m(\tau, -\eta, \zeta, \tau_0) = I_1^{*m}(\tau, \eta, \zeta, \tau_0) \quad (43)$$

Of course, we can select arbitrary starting iterations, for example, by using some approximate analytical, semi-analytical, or numerical solutions of the initial boundary value problem. In this case, the primary atmospheric light scattering and the single surface light reflection have to be estimated exactly. At the same fixed iterative step, the multiple light scattering can be only calculated approximately, but the final resulting radiation field will be estimated sufficiently precise.

Primary light scattering

The exact primary scattering for a vertically nonuniform atmosphere, bounded by a reflecting non-orthotropic surface, can be described by

$$I_1^m(\tau, \eta, \zeta, \tau_0) = \int_0^{\tau_0} \frac{\Lambda(\tau')}{2} \left\{ \int_0^1 p^m(\tau', -\eta, \eta') I_{sur,1}^m(\eta', \zeta, \tau_0) \times \exp\left(\frac{\tau_0 - \tau'}{\eta'}\right) d\eta' + \frac{S}{2} p^m(\tau, \eta, \zeta) \exp\left(-\frac{\tau'}{\zeta}\right) \right\} \times \exp\left(-\frac{\tau - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \quad (44)$$

$$I_1^m(\tau, -\eta, \zeta, \tau_0) = \int_0^{\tau_0} \frac{\Lambda(\tau')}{2} \left\{ \int_0^1 p^m(\tau', \eta, \eta') I_{sur,1}^m(\eta', \zeta, \tau_0) \times \exp\left(-\frac{\tau_0 - \tau'}{\eta'}\right) d\eta' + \frac{S}{2} p^m(\tau, -\eta, \zeta) \exp\left(-\frac{\tau'}{\zeta}\right) \right\} \times \exp\left(-\frac{\tau - \tau'}{\eta}\right) \frac{d\tau'}{\eta} + I_{sur,1}^m(\eta, \zeta, \tau_0) \exp\left(-\frac{\tau_0 - \tau}{\eta}\right) \quad (45)$$

The brightness of an arbitrary non-orthotropic underlying surface for single reflection has the following expression:

$$\eta I_{sur,1}^m(\eta, \zeta, \tau_0) = \frac{S\zeta}{2} y^m(\eta, \zeta) \exp\left(-\frac{\tau_0}{\zeta}\right) + \int_0^1 y^m(\eta, \eta') I_1^{*m}(\tau_0, \eta', \zeta) \eta' d\eta' \quad (46)$$

with $\tau \in [0, \tau_0]$, $\eta \in [0, 1]$, $\zeta \in [0, 1]$ and $m = 0, 1, 2, \dots, M$. In this formula the I_1^{*m} denotes the intensity of primary scattering in a nonuniform atmosphere illuminated by the sun from the side of the top of the atmospheric boundary without the light reflecting from surface placed at the lowest atmosphere boundary. If we substitute Eq. 46 into Eqs. 44 and 45, we obtain

$$I_1^m(\tau, \eta, \zeta, \tau_0) = I_1^{*m}(\tau, \eta, \zeta, \tau_0) + \int_0^1 I_{*1}^m(\tau, \eta, \eta', \tau_0) \zeta \exp\left(-\frac{\tau_0}{\zeta}\right) y^m(\eta', \zeta) \frac{d\eta'}{\eta'} + 2 \int_0^1 I_{*1}^m(\tau, \eta, \eta', \tau_0) \frac{d\eta'}{\eta'} \int_0^1 I_1^{*m}(\tau, \eta'', \zeta, \tau_0) y^m(\eta', \eta'') \eta'' d\eta'' \quad (47)$$

and

$$I_1^m(\tau, -\eta, \zeta, \tau_0) = I_1^{*m}(\tau, -\eta, \zeta, \tau_0) + \int_0^1 I_{*1}^m(\tau, -\eta, \eta', \tau_0) \zeta \exp\left(-\frac{\tau_0}{\zeta}\right) y^m(\eta', \zeta) \frac{d\eta'}{\eta'} + 2 \int_0^1 I_{*1}^m(\tau, -\eta, \eta', \tau_0) \frac{d\eta'}{\eta'} \int_0^1 I_1^{*m}(\tau, \eta'', \zeta, \tau_0) y^m(\eta', \eta'') \eta'' d\eta'' + I_{sur,1}^m(\eta, \zeta, \tau_0) \exp\left(-\frac{\tau_0 - \tau}{\eta}\right) \quad (48)$$

where I_1^m is the brightness in a vertically nonuniform atmosphere illuminated from the sun from the side of the lowest boundary. Neglecting the solar photons that are scattered by the atmosphere and only taking those reflected by the underlying surface one time, that is, only taking into account the direct photons, we can obtain the approximate solution for the intensity $I_1^m(\eta, \pm\zeta, \tau_0)$:

$$I_1^m(\tau, \eta, \zeta, \tau_0) = I_1^{*m}(\tau, \eta, \zeta, \tau_0) + \int_0^1 I_{*1}^m(\tau, \eta, \eta', \tau_0) \zeta \exp\left(-\frac{\tau_0}{\zeta}\right) y^m(\eta', \zeta) \frac{d\eta'}{\eta'} \quad (49)$$

$$I_1^m(\tau, -\eta, \zeta, \tau_0) = I_1^{*m}(\tau, -\eta, \zeta, \tau_0) + \int_0^1 I_{*1}^m(\tau, -\eta, \eta', \tau_0) \zeta \exp\left(-\frac{\tau_0}{\zeta}\right) y^m(\eta', \zeta) \frac{d\eta'}{\eta'} + I_{sur,1}^m(\eta, \zeta, \tau_0) \exp\left(-\frac{\tau_0 - \tau}{\eta}\right) \quad (50)$$

Precise expression of $I_1^{*m}(\tau, \eta, \zeta, \tau_0)$ and $I_1^{*m}(\tau, -\eta, \zeta, \tau_0)$ can be computed with the help of expression Eqs. 23 and 24 using the corresponding exact expression for the source functions that can be obtained taking the sun's position at the top or lowest atmospheric boundaries, that is,

$$B_1^{*m}(\tau, \pm\eta, \zeta, \tau_0) = \frac{S}{4} \Lambda(\tau) p^m(\tau, \pm\eta, \zeta) \exp\left(-\frac{\tau}{\zeta}\right) \quad (51)$$

$$B_{*1}^m(\tau, \mp\eta, \zeta, \tau_0) = \frac{S}{4} \Lambda(\tau) p^m(\tau, \mp\eta, \zeta) \exp\left(-\frac{\tau_0 - \tau}{\zeta}\right) \quad (52)$$

The importance of obtaining precise expression for primary scattering resides also in the need to obtain the grid point to calibrate the radiation fields in environment.

The Gauss-Seidel methods

The application of the Gauss-Seidel method to a multilayer atmosphere bounded from an arbitrary non-orthotropic reflecting surface consists of the discretization of variables τ , η , and φ and the subsequent approximation of the system with a linear algebraic equation without any series expansions (Lenoble, 1985). The atmosphere is divided in $N-1$ layers ($0 = \tau_1 < \tau_2 < \dots < \tau_N = \tau_0$) of optical thickness $\tau = \frac{\tau_0}{N-1}$. For each interval (τ, τ_{i+2}) , Eqs. 9 and 10 may be written in the following form ($\eta > 0$):

$$I(\tau_{i+2}, \eta, \zeta, \varphi - \varphi_0, \tau_0) = I(\tau_i, \eta, \zeta, \varphi - \varphi_0, \tau_0) \exp\left(-\frac{2\Delta\tau}{\eta}\right) + \int_{\tau_i}^{\tau_{i+2}} B(\tau', \eta, \zeta, \varphi - \varphi_0) \exp\left(-\frac{\tau_{i+2} - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \quad i = 1, 2, \dots, N-2 \quad (53)$$

$$I(\tau_i, -\eta, \zeta, \varphi - \varphi_0, \tau_0) = I(\tau_{i+2}, -\eta, \zeta, \varphi - \varphi_0, \tau_0) \exp\left(-\frac{2\Delta\tau}{\eta}\right) + \int_{\tau_i}^{\tau_{i+2}} B(\tau', -\eta, \zeta, \varphi - \varphi_0) \exp\left(-\frac{\tau' - \tau_i}{\eta}\right) \frac{d\tau'}{\eta} \quad i = 1, 2, \dots, N-2 \quad (54)$$

The basic approximation is made according to the following expression for the source function:

$$\int_{\tau_i}^{\tau_{i+2}} B(\tau', -\eta, \zeta, \varphi - \varphi_0, \tau_0) \exp\left(-\frac{\tau_{i+2} - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \approx B(\tau_{i+1}, \eta, \zeta, \varphi - \varphi_0, \tau_0) \left[1 - \exp\left(-\frac{2\Delta\tau}{\eta}\right)\right] \quad (55)$$

while the second approximation is obtained by using the following expression for the angular integral of the source term:

$$\int_{\Omega} \chi(\tau, \omega_j, \omega) I(\tau, \omega, \tau_0) d\omega \approx \sum_{k=1}^M P_k(\tau, \omega_j) I(\tau, \omega, 0) \Delta\omega_k \quad (56)$$

with $d\omega = d\eta d\varphi$. Here, the unit sphere is divided into L domains according the increments $\Delta\omega_j$ of solid angle Ω , where ω_j is the center of $\Delta\omega_j$. The weights $P_k(\tau, \omega_j)$ are defined as integral average values by means of the expression:

$$P_k(\tau, \omega_j) = \frac{1}{\Delta\omega_k} \int_{\Delta\omega_k} \chi(\tau, \omega_j, \omega) d\omega \quad j = 1, 2, 3, \dots, L \quad (57)$$

Making use of these approximations, we obtain the system of linear algebraic equations relative to the unknown values $I(\tau_i, \pm\eta_j, \zeta, \tau_N)$ with ($i = 1, 2, \dots, N; j = 1, 2, \dots, L$). Such a system is solved by the well-known one-step iterative procedure given by Gauss-Seidel's method (Lenoble, 1985). The values of down-going radiation are calculated from one layer to the subsequent layer for each iteration. Namely, on n th iterative step the values of down-going radiation at the $(i+2)$ th layer are obtained by the values of down-going radiation for the i th and $(i+1)$ th layers at the same n th iterative step and by the values of upgoing radiation for the $(i+1)$ th layer at the previous $(n-1)$ th iterative step. At the surface, one takes into account the boundary conditions (7), and then the values of upgoing radiation are calculated from one layer to the following layer upward. The main advantages of this method are:

1. The inner radiation fields are obtained without any additional calculations.
2. The method is simply extended to polarization.
3. The method takes into account the arbitrary dependence of primary optical characteristics of the atmosphere on the height and the arbitrary law of surface reflection.

4. The method with applied point of view makes it possible, in particular, to estimate the role of the nonuniform atmosphere to make the atmospheric corrections of a satellite multispectral data.

The principal disadvantage of the classical Gauss-Seidel method is the linear increase of required computer time with increase of optical thickness of the atmosphere. Therefore, this method is not sufficiently effective for optical thickness $\tau_0 \gg 1$ (Lenoble, 1985). In order to overcome this difficulty, a semi-analytical approach had been proposed to integrate optical thickness $\tau_0 \gg 0$. Until now, we have described the classical Gauss-Seidel method by the three variables τ , η , and φ . We can easily decrease the total number of angular variables using the Fourier series expansions of the phase function $\chi(\tau, \gamma)$ and the bidirectional law of surface reflectance $y(\eta', \eta'', \varphi' - \varphi'')$ according to Fourier series previously defined (Lenoble, 1985). Under these assumptions we get

$$I^m(\tau_{i+2}, \eta, \zeta, \tau_0) = I^m(\tau_i, \eta, \zeta, \tau_0) \exp\left(-\frac{2\Delta\tau}{\eta}\right) + \int_{\tau_i}^{\tau_{i+2}} B^m(\tau', \eta, \zeta, \tau_0) \exp\left(-\frac{\tau_i - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \quad i = 1, 2, \dots, N-2 \quad (58)$$

$$I^m(\tau_i, -\eta, \zeta, \tau_0) = I^m(\tau_{i+2}, -\eta, \zeta, \tau_0) \exp\left(-\frac{2\Delta\tau}{\eta}\right) + \int_{\tau_i}^{\tau_{i+2}} B^m(\tau', -\eta, \zeta, \tau_0) \exp\left(-\frac{\tau' - \tau_i}{\eta}\right) \frac{d\tau'}{\eta} \quad i = 1, 2, \dots, N-1 \quad (59)$$

Thus, the Gauss-Seidel method is now given by this approximation:

$$\int_{\tau_i}^{\tau_{i+2}} B^m(\tau_i, \eta, \zeta, \tau_0) \exp\left(-\frac{\tau_{i+2} - \tau'}{\eta}\right) \frac{d\tau'}{\eta} \approx B^m(\tau_{i+1}, \eta, \zeta, \tau_0) \left[1 - \exp\left(-\frac{2\Delta\tau}{\eta}\right)\right] \quad (60)$$

The second step is now to write the form of classical Gauss quadratures:

$$\int_0^1 f(\eta) d\eta \approx \sum_{j=1}^L f(\eta_j) w_j \quad (61)$$

Then, the application of the Gauss-Seidel method for the radiation field azimuth's harmonics is reduced to the basic boundary value problem for each $m = 0, 1, \dots, M$ of the following system of linear algebraic equations to unknown values $I^m(\tau_i, \eta_j, \zeta, \tau_0)$ with $i = 1, 2, \dots, N-2$ and $j = 1, 2, \dots, L$:

$$I^m(\tau_{i+2}, \eta_j, \zeta, \tau_0) = I^m(\tau_i, \eta_j, \zeta, \tau_0) \exp\left(-\frac{2\Delta\tau}{\eta_j}\right) + B^m(\tau_{i+1}, \eta_j, \zeta, \tau_0) \left[1 - \exp\left(-\frac{2\Delta\tau}{\eta_j}\right)\right] \quad (62)$$

$$I^m(\tau_i, -\eta_j, \zeta, \tau_0) = I^m(\tau_{i+2}, -\eta_j, \zeta, \tau_0) \exp\left(-\frac{2\Delta\tau}{\eta_j}\right) + B^m(\tau_{i+1}, -\eta_j, \zeta, \tau_0) \left[1 - \exp\left(-\frac{2\Delta\tau}{\eta_j}\right)\right] \quad (63)$$

$$\eta_j I^m(\tau_N, -\eta_j, \zeta, \tau_0) = \sum_{k=1}^L y^m(\eta_j, \eta_k) I^m(\tau_N, \eta_k, \zeta, \tau_0) \eta_k \omega_k + \frac{S}{2} y^m(\eta_j, \zeta) \exp\left(-\frac{\tau_N}{\zeta}\right) \quad (64)$$

where

$$B^m(\tau_i, \eta_j, \zeta, \tau_0) = \frac{S}{4} \Lambda(\tau_i) p^m(\tau_i, \eta_j, \zeta) \exp\left(-\frac{\tau_i}{\zeta}\right) + \frac{\Lambda(\tau_i)}{2} \sum_{k=1}^L \omega_k \left[p^m(\tau_j, \eta_j, \eta_k) I^m(\tau_i, \eta_k, \zeta, \tau_0) + p^m(\tau_i, -\eta_j, \eta_k) I^m(\tau_i, -\eta_k, \zeta, \tau_0) \right] \quad (65)$$

$$B^m(\tau_i, -\eta_j, \zeta, \tau_0) = \frac{S}{4} \Lambda(\tau_i) p^m(\tau_i, -\eta_j, \zeta) \exp\left(-\frac{\tau_i}{\zeta}\right) + \frac{\Lambda(\tau_i)}{2} \sum_{k=1}^L \omega_k \left[p^m(\tau_j, -\eta_j, \eta_k) I^m(\tau_i, \eta_k, \zeta, \tau_0) + p^m(\tau_i, \eta_j, \eta_k) I^m(\tau_i, -\eta_k, \zeta, \tau_0) \right] \quad (66)$$

It should be noted that now we have only two variables, τ and η , while the values ζ and τ_0 are parameters. This system is solved by means of the direct iterative procedure analogously to the previously defined iterative schemes:

$$I_{n+1}^m(\tau_{i+2}, \eta_j, \zeta, \tau_i) = I_n^m(\tau_i, \eta_j, \zeta, \tau_0) \exp\left(-\frac{2\Delta\tau}{\eta_j}\right) + B_n^m(\tau_{i+1}, \eta_j, \zeta, \tau_0) \left[1 - \exp\left(-\frac{2\Delta\tau}{\eta_j}\right)\right] \quad (67)$$

$$I_{n+1}^m(\tau_i, -\eta_j, \zeta, \tau_i) = I_n^m(\tau_{i+2}, -\eta_j, \zeta, \tau_0) \exp\left(-\frac{2\Delta\tau}{\eta_j}\right) + B_n^m(\tau_{i+1}, -\eta_j, \zeta, \tau_0) \left[1 - \exp\left(-\frac{2\Delta\tau}{\eta_j}\right)\right] \quad (68)$$

$$\eta_j I_{n+1}^m(\tau_N, -\eta_j, \zeta, \tau_0) = \sum_{k=1}^L y^m(\eta_j, \eta_k) I_n^m(\tau_N, \eta_k, \zeta, \tau_0) \eta_k \omega_k + \frac{S}{2} y^m(\eta_j, \zeta) \exp\left(-\frac{\tau_N}{\zeta}\right) \quad (69)$$

with $n = 0, 1, \dots, K; i = 1, 2, \dots, N-2; j = 1, 2, \dots, L$.

The azimuth harmonics of the source function are given by

$$B_n^m(\tau_i, \eta_j, \zeta, \tau_0) = \frac{S}{4} \Lambda(\tau_i) p^m(\tau_i, \eta_j, \zeta) \exp\left(-\frac{\tau_i}{\zeta}\right) + \frac{\Lambda(\tau_i)}{2} \sum_{k=1}^L w_k [p^m(\tau_i, \eta_j, \eta_k) I_n^m(\tau_i, \eta_k, \zeta, \tau_0) + p^m(\tau_i, -\eta_j, \eta_k) I_n^m(\tau_i, -\eta_k, \zeta, \tau_0)] \quad (70)$$

$$B_n^m(\tau_i, -\eta_j, \zeta, \tau_0) = \frac{S}{4} \Lambda(\tau_i) p^m(\tau_i, -\eta_j, \zeta) \exp\left(-\frac{\tau_i}{\zeta}\right) + \frac{\Lambda(\tau_i)}{2} \sum_{k=1}^L w_k [p^m(\tau_i, -\eta_j, \eta_k) I_n^m(\tau_i, \eta_k, \zeta, \tau_0) + p^m(\tau_i, \eta_j, \eta_k) I_n^m(\tau_i, -\eta_k, \zeta, \tau_0)] \quad (71)$$

Iterative process is over when the following conditions are satisfied:

$$\max \left| \frac{I_{n+1}^m(\tau_i, \pm \eta_j, \zeta, \tau_0) - I_n^m(\tau_i, \pm \eta_j, \zeta, \tau_0)}{I_n^m(\tau_i, \pm \eta_j, \zeta, \tau_0)} \right| \leq \epsilon \quad (72)$$

$i = 1, \dots, N; \quad j = 1, \dots, L$

In such a way, the classical Gauss-Seidel method has been modified by the Fourier expansions of the phase functions and by the law of surface reflectance, which reduce the total number of variables. Finally, let us note that the values of intensity in the horizontal direction $\eta = 0$ may be easily calculated by the intensities at Gaussian points obtained above. Namely, at $\eta = 0$ we get for any optical depth τ :

$$I^m(\tau_i, 0, \zeta, \tau_0) = \frac{\Lambda(\tau)}{2} \int_{-1}^1 p^m(\tau, 0, \eta') I^m(\tau, \eta', \zeta, \tau_0) d\eta' + \frac{S}{4} \Lambda(\tau) p^m(\tau, 0, \zeta) \exp\left(-\frac{\tau}{\zeta}\right) \quad (73)$$

$$\approx \sum_{k=1}^L \omega_k p^m(\tau, 0, \eta_k) [I^m(\tau, \eta_k, \zeta, \tau_0)] + I^m(\tau, -\eta_k, \zeta, \tau_0)] + \frac{S}{4} \Lambda(\tau) p^m(\tau, 0, \zeta) \exp\left(-\frac{\tau}{\zeta}\right) \quad (74)$$

The input optical models for the earth's atmosphere

One important piece of information comes from the models of atmospheric aerosol that enter into the mathematics and algorithms previously described. The World Climate Programme (1982) selected a certain number of aerosol models with their optical properties. Levoni et al. (1997) expanded the number of aerosols and clouds selected by WMP, giving all the optical properties and the algorithms to calculate them and also including the computation to obtain the Legendre's polynomials.

Of course, the atmospheric phase functions have to be normalized according to the already-mentioned rules, and $\chi(\tau, \eta, \zeta, \varphi - \varphi_0)$ and $\chi^m(\tau, \eta, \zeta)$, and must be placed under the condition of reciprocal angular symmetry:

$$\chi(\tau, \eta, \zeta, \varphi - \varphi_0) = \chi(\tau, \eta, \varphi, -\varphi_0) \quad (75)$$

$$\chi^m(\tau, \eta, \zeta) = \chi^m(\tau, \zeta, \eta) \quad (76)$$

The parameterization and approximation of atmospheric phase function can be done, for example, with the help of some of following imitative models:

$$\chi(\tau, \gamma) \cong \frac{1 - g^2(\tau)}{(1 + g^2(\tau) - 2g(\tau)\cos\gamma)^{\frac{3}{2}}} |g(\tau)| < 1 \quad (77)$$

$$\chi(\tau, \gamma) \cong \frac{4\alpha g(\tau)(1 - g^2(\tau))^{2\alpha}}{r \left[(1 + g^2(\tau))^{2\alpha} - (1 - g^2(\tau))^{2\alpha} \right] (1 + g^2(\tau) - 2g(\tau)\cos\gamma)^{\alpha+1}} \quad (78)$$

$\alpha > -\frac{1}{2}, \quad |g(\tau)| < 1$

$$\chi(\tau, \gamma) \cong \alpha \frac{1 - g_1^2(\tau)}{(1 + g_1^2(\tau) - 2g_1(\tau)\cos\gamma)^{\frac{3}{2}}} + (1 - \alpha) \frac{1 - g_2^2(\tau)}{(1 + g_2^2(\tau) - 2g_2(\tau)\cos\gamma)^{\frac{3}{2}}} \quad (79)$$

$|g_1(\tau)| < 1 \quad |g_2(\tau)| < 1$

$$\chi(\tau, \gamma) \cong a \frac{1 - g^2(\tau)}{(1 + g^2(\tau) - 2g(\tau)\cos\gamma)^{\frac{3}{2}}} + (1 - a) \sum_{i=0}^{\tilde{M}} x_i(\tau) P_n(\cos\gamma) \quad (80)$$

$$|g(\tau)| < 1; \quad |x_i(\tau)| < 2i + 1; \quad \tilde{M} < M$$

where g is the Heney-Greenstein parameter. The problem of the approximation of atmospheric phase function $\chi(\tau, \gamma)$ with the help of parameterized phase functions $\tilde{\chi}(\tau, \gamma)$ is a typical problem of optimal choosing

of parameters $g(\tau)$, $g_1(\tau)$, $g_2(\tau)$, a , $x_i(\tau)$, and α according to the well-known condition of the optimization:

$$\min_{(i,j)} \left| \sum_{i=0}^K \sum_{j=0}^L \chi(\tau_j, \gamma_i) - \tilde{\chi}(\tau_j, \gamma_i, g, g_1, g_2 \dots) \right|^2 = \varepsilon^2 \quad (81)$$

where ε is the level of error. It should be remarked that the principal disadvantages arising for the Fourier harmonics $p^m(\eta, \zeta)$ are the non-stability of numerical schemes and the accuracy loss for calculated values in the case of a large numbers $m \gg 1$.

We can now define our strategy on the level of the informational contents of input optical models by the following statements:

1. The input phase function is given according to initial grid point $G(i, j, \dots k)$ and error level of ε . This is the first level of informational content that is determined by the calculation according to Mie theory (Levoni et al., 1997).
2. The problems of the determination of structured coefficients x_n and polynomials functions p^m in the points γ_i of the chosen initial grid are performed by the determination of unknown spectra for known function $\chi(\gamma)$ given in the abovementioned grid point only. It is well known, from spectral theory of signals, that discretization levels cannot be arbitrary.

According to the Nyquist-Kotelnikov theorem, the current frequency of discretization L should exceed at least twice that of the high signal frequency M :

$$L \geq 2M \quad (82)$$

However, we emphasize that high discretization levels of input phase function $\chi(\tau_l, \gamma_i)$, needed for stable and reliable determination of their high Fourier azimuth's harmonics, cannot be achieved practically. Therefore, for numerical interpolations of phase function $\chi(\tau_l, \gamma_i)$ between knots of initial grid points, taking into account the interpolation's accuracy $\tilde{\varepsilon}$, the determination of x_n and p^m provides a good numerical approach. As a result, we obtain a second level of information that is interpolated with a new grid point $\tilde{g}$ and a second level of error ε^* . According to recommendations given, the spline interpolation could provide good computational accuracy for interpolated phase function.

3. A third information content level arises after summing structured coefficients x_n and function p^m in the knots of interpolated grid points, and the determination of its total number N and M according to conditions of minimization is needed:

$$\sum_{j=0}^L \left| \chi_{\tilde{g}}(\tau_l, \gamma_i) - \sum_{i=0}^M x_i P_i(\cos \gamma_j) \right|^2 = \sigma^2 = \min \quad (83)$$

$$\sum_{j=0}^L \left| \chi_{\tilde{g}}(\tau_l, \gamma_i) - \sum_{i=0}^M p_{\tilde{g}}^m(\eta_i, \zeta_i, \tau_l) \cos m \varphi_k \right|^2 = \sigma^2 = \min \quad (84)$$

It must be clear that the previous sums will be carried out only after calibration procedures for coefficients x_n and function p^m . Naturally, the procedures mentioned cannot change the grid points $\tilde{g}$ but only correct values of x_n and p^m .

4. Finally, the final informational level arises after the compression of total numbers of coefficients x_n and azimuth's harmonics p^m on the basis of new approaches connected with mutual estimation of azimuth harmonics radiation fields of primary atmospheric scattering and single-bottom reflection from one side and multiple scattered and reflected radiation field from another side.

This contribution is increased when the number of harmonics m grows. This conclusion is a serious foundation of the compression of total number of azimuth's harmonics in the sum of Eq. 84, according to the following statement:

$$\frac{\Delta p^m(\eta, \zeta, \tau_0)}{p_{(1)}^m(\eta, \zeta, \tau_0)} < \varepsilon \quad m \Rightarrow m_{cut} \quad (85)$$

Further, the parameter L determines the accuracy of calibrated coefficients and functions, while the parameter M influences the approximation's accuracy determined with the help of series Eq. 84. According to Nyquist-Kotelnikov's theorem, both parameters are subjected to relation Eq. 82.

This estimation, in accordance with Nyquist-Kotelnikov's theorem, shows clearly that initial total number of radiation fields azimuth's harmonics M can be compressed significantly in the frame of accuracy estimation given by Eq. 85 lossless of general information content for calculating radiation intensities. Here, the unique parameter needed to approximate the input atmospheric phase function $\chi(\tau, \gamma)$ with help of coefficients x_n and functions $p^m(\eta, \zeta)$, taking into account beforehand approximation accuracy ε , is M :

$$\max_{(\tau_l, \gamma_s)} \left| \frac{\chi(\tau_l, \gamma_s) - \sum_{n=0}^M x_n(\tau_l) P_n(\cos \gamma_s)}{\chi(\gamma_s)} \right| = \tilde{\varepsilon}_1 \quad (86)$$

$$\max_{(\gamma_i, \eta_i, \zeta_j, \varphi_k)} \left| \frac{\chi_{\tilde{g}}(\tau_l, \eta_i, \zeta_j, \varphi_k) - \sum_{m=0}^M p_{\tilde{g}}^m(\tau_e, \eta_i, \zeta_j) \cos m \varphi_k}{\chi_{\tilde{g}}(\gamma_i, \eta_i, \zeta_j, \varphi_k)} \right| = \tilde{\varepsilon}_2 \quad (87)$$

The input optical models for the ground surface

Analogous considerations for the input optical models of the atmosphere may be carried out for input optical models of the ground surface. However, in the case of underlying

surfaces, there are some differences. Atmospheric optical models connect directly chemical and physical parameters of atmospheric aerosol's particles with appropriate atmospheric optical parameters on the basis of Mie theory independently from their influence of natural underlying surfaces. The ideal for surface optical models should be to use appropriate surface diffraction theory. Unfortunately, in real situations the correct approach is replaced by measurements or imitative modeling linking the scattering and absorbing atmosphere. In this case, the final determination of surface optical parameters will significantly depend on the atmospheric optical state. Publications, dedicated to analysis of empirical, semiempirical, and theoretical data, show that differently from the atmosphere the problem of representative and classified optical models of ground surface have not been yet fully solved. Nevertheless, we will take into account real situations with proper parameterization models applied to empirical and semiempirical data. Most representative optical models of the ground surface can be described using the bidirectional reflectance. However, in this case, some problems concerning the bidirectional surface reflectance $r_{sur}(\eta, \zeta, \varphi)$ arise. This is determined as ratio of reflected radiation to the incident flux under the same conditions, with corresponding surface phase function $y(\eta, \zeta, \varphi - \varphi_0)$. Usually, empirical and semiempirical values of a surface bidirectional reflectance are represented by

$$r_{sur} = A(\zeta)\chi_{sur}(\eta, \zeta, \varphi - \varphi_0) \quad (88)$$

where $A(\zeta)$ is the albedo of the surface. By simple considerations we can establish obvious relation between surface phase function $\chi_{sur}(\eta, \zeta, \varphi)$ and probabilistic function $y(\eta, \zeta, \varphi)$, making use of the normalized condition:

$$y(\eta, \zeta, \varphi) = 2\eta\chi_{sur}(\eta, \zeta, \varphi) \quad (89)$$

In order to represent the probabilistic function $y(\eta, \zeta, \varphi)$, the following models can be used as a basic parameterization model for numerical approximations of empirical and semiempirical surface bidirectional brightness coefficients:

1. Orthotropic case

$$y(\eta, \zeta, \varphi) = 2\eta A(\zeta) \quad (90)$$

2. Lambertian surface

$$y(\eta, \zeta, \varphi) = 2\eta A \quad (91)$$

3. Minnaert model

$$y(\eta, \zeta, \varphi) = C\eta^q\zeta^{q-1} \quad (92)$$

4. Mirror surface

$$y(\eta, \zeta, \varphi) = 2\pi\delta(\eta - \zeta)\delta(\varphi)r(\eta) \quad (93)$$

Where $r(\eta)$ is Fresnel's coefficient of mirror reflection

5. Panfilov model (cited by Kondratiev et al., 1992)

$$y(\eta, \zeta, \varphi) = \frac{2\eta d}{1+d} \left[1 + \frac{1-g^2}{d(1+g^2-2g\cos\gamma)^{\frac{3}{2}}} \right] \quad (94)$$

Where g is known as the Heney-Greenstein parameter and d depends on the kind of underlying surface

6. Pinty-Dickinson-Verstraete model (Verstraete et al., 1990) in which bidirectional reflectance of a canopy normalized with respect to the reflectance of a perfectly reflecting Lambertian surface under the same conditions of the sun's illumination and observation is equal to:

$$r_{sur} = \frac{\Lambda_{sur}}{4} \frac{k_1(\zeta)\zeta}{k_1(\zeta)\eta + k_2(\eta)\zeta} \left[P(\pi - \gamma)P_{rel}(\pi - \gamma) + H\left(\frac{\zeta}{k_1(\zeta)}\right)H\left(\frac{\eta}{k_1(\eta)}\right) - 1 \right] \quad (95)$$

where the following symbols are used:

$$k_1(\zeta) = \Phi_1 + \Phi_{2\zeta} \quad (96)$$

$$k_2(\eta) = \Phi_1 + \Phi_{2\eta}$$

$$\Phi_1 = 0.5 - 0.6\chi - 0.033\chi^2 \quad (97)$$

$$\Phi_2 = 0.877(1 - 2\Phi_1) \quad (98)$$

The semiempirical parameter χ is varied in the interval $[-0.4, 0.6]$, accepting the following magnitudes for real canopy's kind:

- (a) 0 for a canopy with uniform leaf orientation distribution (equal probability for all leaf orientations).
- (b) 0.6 for a planophile canopy (mostly horizontal leaves).
- (c) -0.4 for an erectophile canopy (mostly vertical leaves).

The modeling auxiliary function $P_{rel}(\pi - \gamma)$ is given by:

$$P_{rel}(\pi - \gamma) = 1 + \frac{1}{1 + F(\pi - \gamma)} \quad (99)$$

where

$$F(\pi - \gamma) = 4 \left(1 - \frac{4}{3\pi} \right) \frac{\eta}{k_2(\eta)} G(\eta, \zeta, \varphi) \quad (100)$$

and

$$G(\eta, \zeta, \varphi) = \frac{1}{2\rho\Lambda} \left[\frac{1-\zeta^2}{\zeta^2} + \frac{1-\eta^2}{\eta^2} + 2 \frac{\sqrt{(1-\zeta^2)(1-\eta^2)}}{\eta\zeta} \cos(\varphi - \varphi_0) \right] \quad (101)$$

The empirical parameters ρ varied in interval $0.01 \leq \rho \leq 2$. The function $P(\pi - \gamma)$ represents the modeling surface phase function for given semiempirical model by using, for example, the known Heney-Greenstein phase function of Eq. 80 for $\tilde{\gamma} = \pi\gamma$.

Now let us represent the bidirectional reflectance given by Eq. 61 with a help from Eq. 62:

$$2\eta r_{sur}(\eta, \zeta, \varphi) = A(\zeta)y(\eta, \zeta, \varphi) \quad (102)$$

Taking into account normalized condition 3, we obtain the following correlation:

$$y(\eta, \zeta, \varphi) = \frac{4\pi\eta r_{sur}(\eta, \zeta, \varphi)}{\int_0^{2\pi} d\varphi \int_0^1 r_{sur}(\eta, \zeta, \varphi)\eta d\eta} \quad (103)$$

It should be noted that replacing surface phase function $\chi_{sur}(\eta, \zeta, \varphi)$ with the appropriate probabilistic functions $y(\eta, \zeta, \varphi)$ is not quite correct because the transformation, although having some advantages, simultaneously violates the fundamental physical properties of reciprocal angular symmetry $\chi_{sur}(\eta, \zeta, \varphi) = \chi_{sur}(\zeta, \eta, \varphi)$. Thus, the input optical models for the atmosphere we can enter into the analysis of appropriate informational levels are connected with the parameterization and structuring of surface optical quantities. The parameterization of the bidirectional surface reflectance is given the minimization's condition:

$$\min_{i,j,k} \sum_i \sum_j \sum_k |y(\eta_i, \zeta_j, \varphi_k) - \tilde{y}(\eta_i, \zeta_j, \varphi_k)|^2 = \sigma^2 \quad (104)$$

Afterward, the previous procedures of structuring, due to the use of Legendre's polynomials series expansion and Fourier series expansion, can be repeated for the discussed surface case without any great difficulty. The same remarks hold for the procedure of calibration and the sum of structured elements. The main point is the necessity of connecting the total number of the coefficients and functions calculated separately for the atmosphere and underlying surfaces. Thus, as results we have input a unified optical model for an atmosphere-surface system with four levels of informational content, which is needed to accurately define successive radiation fields modeling.

Discrete ordinate method

The discrete ordinate method isolates the azimuth dependence of Eqs. 5 and 6, expanding the phase function into the Legendre polynomial. Using the addition theorem for spherical harmonics and taking into account the orthogonality properties of the Legendre polynomials, we obtain that the phase function can be written as:

$$\chi(\tau, \gamma) = \sum_{m=0}^{2N-1} (2 - \delta_{0m}) p^m(\tau, \eta, \zeta) \cos m(\varphi - \varphi_0) \quad (105)$$

where p^m is given by Eq. 15 and δ_{0m} is the Kronecker delta where $\delta_{0m} = 1$ for $m = 0$ and $\delta_{0m} = 0$ otherwise. Expanding the radiation in the similar way, we obtain:

$$I(\tau, \eta, \zeta, \varphi - \varphi_0) = \sum_{m=0}^{2N-1} I^m(\tau, \eta) \cos m(\varphi - \varphi_0) \quad (106)$$

The procedure is similar to that described in the Fourier expansion section.

The application to the multilayer atmosphere-surface system requires having sufficiently large layers to resolve the radiation in terms of its angular dependence and its optical thickness. The discrete ordinate method was developed by K. Stamnes and Swanson (1981), Stamnes (1982), and Stamnes et al. (1988). Thomas and Stamnes (1999) extended the approach to an atmosphere-ocean system, choosing N_1 layers for the atmosphere and N_2 layers for the ocean. In such a case, Eqs. 19 and 20 can be solved approximating the integral over polar angles by a quadrature sum consisting of $2N_1$ terms in the atmosphere, N_1 terms in the upper hemisphere, and N_1 terms in the downward hemisphere. These last terms are refracted through the interface when the radiation penetrates into the water. Of course, in order to represent the total reflection, we need more terms that can be chosen, such as $2N_2$, ($N_2 > N_1$) for the radiation in water, yielding $2(N_2 - N_1)$ terms of the total reflecting region.

Thus, the solution of the discrete ordinate method applied to the atmosphere (superscripted with a) can be written for the p th layers as:

$$I_p^{\pm}(\tau, \pm\eta_i^a) = \sum_{j=1}^{N_1} \left[C_{-jp} g_{-jp}^a(\pm\eta_i^a) \exp(k_{jp}^a \tau) + C_{jp} g_{jp}^a(\pm\eta_i^a) \exp(-k_{jp}^a \tau) \right] + U_p(\tau, \pm\eta_i^a) \quad (107)$$

where $i = 1, \dots, N_1$ and $p \leq N_1$. The same is true for the p th layers in water (superscripted with w):

$$I_p^{\pm}(\tau, \pm\eta_i^w) = \sum_{j=1}^{N_1} \left[C_{-jp} g_{-jp}^w(\pm\eta_i^w) \exp(k_{jp}^w \tau) + C_{jp} g_{jp}^w(\pm\eta_i^w) \exp(-k_{jp}^w \tau) \right] + U_p(\tau, \pm\eta_i^w) \quad (108)$$

where $i = 1, \dots, N_2$ and $N_1 < p \leq N_1 + N_2$. The eigenvalues k_{jp}^a, k_{jp}^w and the eigenvectors g_{jp}^a, g_{jp}^w are determined solving the homogeneous associate of Eq. 19 with the source terms put equal to zero. The term $U_p^m(\tau, \pm\eta_i)$ is the particular solution.

The coefficients $C_{\pm jp}$ are determined by:

1. The boundary conditions at the top of the atmosphere and at the bottom of the ocean

2. Radiation continuity conditions at each interface between the layers
3. Fresnel's equations at the atmosphere water interface

This leads to a system of linear algebraic equations:

$$\mathbf{AX} = \mathbf{B} \quad (109)$$

whose solution is $\mathbf{X} = \mathbf{A}^{-1} \mathbf{B}$. Here, $\mathbf{A}$ contains information about the properties of the optical system and the bidirectional reflectance distribution function (BRDF). The order of $\mathbf{A}$ is defined by the unknown coefficient $C_{\pm jp}$. The column vector $\mathbf{X}$ contains information on $C_{\pm jp}$, the matrix $\mathbf{B}$ contains information about the particular solutions as well as the lower boundary emissivity.

Since the phase function is the sum of Rayleigh (molecular) scattering plus the Mie scattering due to the effect of aerosol in atmosphere, an improved solution was proposed by T. Nakajima and Tanaka (1981), which computes the Rayleigh scattering exactly. The theoretical approach can be found in Levoni et al. (2001). This approach improves the accuracy of the solution and reduces the computational time, which largely depends on the number of momenta used for Lagrange polynomials.

Spurr (2001) has developed a solution based on the recognition that if the integral properties of one of the layers in a system are perturbed, there is no need to repeat the entire computation from scratch. By its analysis, he deduced that the solution of the problem can be defined by a matrix form of the type:

$$\mathbf{AX}' = \mathbf{B}' \quad (110)$$

In this relation, which is similar to the previous one Eq. 109, the vector column $\mathbf{B}$ depends on the layer parameters that have been changed. Since the matrix $\mathbf{A}$, unperturbed, is solved by LU decomposition, the solution of Eq. 110 can be found by back-substitution using the LU factorized form of $\mathbf{A}$. The Spurr model, named LIDORT, also contains perturbations in layer thermal emission as well as surface albedo.

The first radiative transfer model, LOWTRAN (Kneizys et al., 1988), and the successive version, MODTRAN (Berk et al., 1989), use the discrete ordinate methods, even with a limited number of momenta of Legendre polynomials. The success of the discrete ordinate methods is due to their flexibility, at least in the frame of remote sensing, so various other authors have adopted this approach, dressing the solutions with the optical features of the atmosphere and the surface to obtain a suitable model. These include the SBDART (Ricchiazzi et al., 2007), libRadtran (Mayer et al., 2005), and, in the IR, HARTCODE (Miskolczi et al., 1998).

Statistical modeling or Monte Carlo methods

Monte Carlo methods provide approximate solutions to the radiative transfer equation by sampling the trajectories

of a large number of photons. There are at least three Monte Carlo methods:

1. Forward with bin averaging, where the photons are averaged over various chosen ranges of solid angles.
2. Forward with discrete angles, where the radiation is estimated at each collision for particular predetermined angles.
3. Backward or adjoint with discrete angles, where the photons start at particular angles at the sensor and retrace back the path. Such a method is the most useful for spherical atmospheres.

In solar radiation problems, the photons are introduced with random initial position at the boundary of the domain. Their directions are chosen according to the illumination conditions. The photons then begin their flight. In an efficient Monte Carlo program, a statistical weight is associated with each photon. In practice, a sample transmission is chosen uniformly between 0 and 1. The photon travels along its path until the cumulative extinction along the photon's path reaches the value that corresponds to this transmission. The photon experiences a scattering event in which the scattering angle is determined by choosing a random number and comparing this to the cumulative phase function. After a photon's new direction is determined with a new transmission, this process repeats until the photon leaves the domain or is absorbed. During scattering the absorption is normally modeled by multiplying the photon weight (initially 1) by the single scattering albedo of the medium doing the scattering. At each collision this weight is multiplied by the ratio of the total scattering cross section to the total cross section for all processes, corrected for the absorption probability. The photon trajectory is terminated when the statistical weight falls below a preassigned value (usually 10^{-5}). In highly variable media, a technique called maximum cross section is often used to reduce the per-photon variance of flux. The cumulative extinction at each step is scaled by the maximum extinction within the domain, and the likelihood of scattering at the end of each step depends on the ratio of the local to the maximum extinction. Radiative quantities are computed by these weighted photons.

The atmosphere may contain more than one optically active component (e.g., aerosols, gases, and clouds often coexist). The extinction accumulated along the photon trajectory is the extinction due to all components. Scattering may be treated either by averaging the single scattering phase functions (weighted by the extinction and single scattering albedo) or by choosing the phase function for a single component based on the relative amounts of extinction.

A technique called Russian roulette can help to speed up calculations in absorbing media. At each scattering event, the photon weight is compared with a random number. If the weight exceeds the random number, the photon trajectory is terminated, otherwise the photon weight is reset to 1. In Cornish-Gilliflower release, for instance (<http://code.google.com/p/i3rc-monte-carlo-model/>), the

implementation Russian roulette is performed only when the photon weight is less than one-half. Intensity is computed using a technique called local estimation. At each scattering event, the total extinction to the boundary from the scattering location is computed in each direction at which intensity is desired. The intensity at the exiting location is then incremented by the product of the transmission along that direction, the phase function of the scattering particle evaluated at the angle between the direction the photon is traveling and the intensity direction, and the photon weight. The phase functions of large particles (i.e., those much larger than the wavelength of light being simulated) have large forward peaks due to diffraction. These peaks may be orders of magnitude larger than the phase function, only a few degrees off-axis. If the forward peak happens to align with one of the directions at which intensity is being computed, the contribution from an individual scattering event can be very large, leading to large variance between calculations (and hence large error estimates). In order to reduce this variance, the phase function used for local estimation can be replaced by a hybrid that replaces the original phase function at small angles by a Gaussian of user-specified width. The hybrid phase function is constructed to be continuous and properly normalized.

From 1D to 3D Monte Carlo

Let us now apply the previous concepts to a homogeneous atmosphere with extinction coefficient $\beta(\mathbf{r})$ and phase function $P(\mathbf{r}, \omega \cdot \omega')$. One can calculate the 1D Monte Carlo radiative transfer using a random number generator (numerical algorithm) to produce the random numbers, r_n , between 0 and 1 with a probability distribution function $pdf = 1$. From r_n , one can generate another set of random numbers as $r_x = -\ln(r_n)$ with $pdf = \exp(-r_x)$ and r_x between 0 and ∞ . Then, in order to determine the trajectory of the photons, one can:

1. Determine the starting position of a photon given by $\mathbf{x}_0$ with direction η, ϕ .
2. Generate a photon path length using random numbers $x = r_x$ using the random numbers r_x .
3. Calculate a new photon position $\mathbf{x}_0 + \mathbf{x}$ called the event point (or collision point).
4. Analyze what can happen with the photon at this event point by generating the random number r_n and comparing it with the single scattering albedo Λ . If $r_n > \Lambda$, then the photon is absorbed and the cycle returns to point 1 for a new photon. If $r_n < \Lambda$, the photon is scattered and the process continue to the next step.
5. Find a new direction for the scattered photon using the phase function to calculate the cumulative probability function to relate the scattering angle to a random number.
6. Repeat, starting from the point 3, until all photons are analyzed.

Let us consider now the inhomogeneous atmosphere. We can split it into a homogeneous grid. The point of interaction (or event point) $r_x = \sum \Delta_j \Delta_j \beta(r)$ where $\Delta_j =$ step size in each grid and the free path length is $l = \frac{1}{\beta(r)}$.

In order to estimate the accuracy of the method with respect to the number of photons, one needs to estimate the Monte Carlo error given by the expectation and variance of random variables. They can be computed by the obtained number N of independent values $x_i (i = 1, 2, 3 \dots N)$ of a random variable r_n , each weighted by the probability with which the value is taken. A detailed analysis is described in Marshak and Davis (2005).

In order to obtain a 3D radiative transfer equation, one needs to reconvert the plane-parallel radiative transfer Eqs. 5 and 6 into one equation, taking into account three spatial dimension (x, y , and z) and two angular dimensions (θ, ϕ). This procedure is different than the one we have used in the plane-parallel system previously described, where the azimuth angles and zenith were, respectively, represented by the Fourier series and discrete ordinates. Since in 3D radiative transfer the spatial dimensions cannot be solved analytically, as the optical depth is in plane-parallel transfer, they must be solved in other way. One way is to use the so-called spherical harmonics discrete ordinate method (SHDOM) developed by Pincus and Evans (2009). We do not go into such a method because our purpose here is to describe the solution for the Monte Carlo method.

Let us now assume a cylinder to be infinitesimally small and located somewhere in the atmosphere. It is oriented along an axis ω that is not pointing directly at the target. Rewrite now the integrodifferential radiative transfer radiation Eq. 5 for an unpolarized beam that describes the change of radiation $I(\mathbf{r}, \omega)$ into direction ω as

$$\omega \nabla I(\mathbf{r}, \omega) = -\beta(\mathbf{r})I(\mathbf{r}, \omega) + \frac{\beta_s(\mathbf{r}, \omega)}{4\pi} \int_{4\pi} I(\mathbf{r}, \omega') P(\mathbf{r}, \omega \cdot \omega') d\omega' + S(\mathbf{r}, \omega) \quad (111)$$

where ω is the direction vector with component $(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ and θ and ϕ are zenith and azimuth angles; $\beta(\mathbf{r})$ is the extinction coefficients; and $\beta_s(\mathbf{r})$ is the scattering coefficient.

Let us rearrange this equation into an integral equation:

$$\left(1 + \frac{1}{\beta(\mathbf{r})} \omega \nabla\right) I(\mathbf{r}, \omega) = -\frac{\beta_s(\mathbf{r})}{\beta(\mathbf{r})} \frac{1}{4\pi} \int_{4\pi} I(\mathbf{r}, \omega') P(\mathbf{r}, \omega \cdot \omega') d\omega' + S(\mathbf{r}, \omega) \frac{1}{\beta(\mathbf{r})} \quad (112)$$

where the ratio $\frac{\beta_s(\mathbf{r})}{\beta(\mathbf{r})}$ is the single scattering albedo. Such integral equation can be solved with the aid of Green's function. In fact, let us transform it into a simple form:

$$L_x f(x) = u(x) \quad (113)$$

where

$$L_x = \left(1 + \frac{1}{\alpha(x)} \frac{d}{dx}\right)$$

is the operator of the left part of the equation, $f(x)$ is the radiance, and $u(x)$ is the right part of Eq. 112, continuous function in the D domain selected. The extinction coefficient $\beta(\mathbf{r})$ has been replaced by $\alpha(x)$. In order to solve the Eq. 113, we look for a function $g: C^1(D) \rightarrow C(D)$ such that $\tilde{L}(g(h)) = h$ where $y(t) = g(h(t))$. This is a convolution equation of the form $y = g \otimes h$ so that the solution is

$$y(t) = \int_a^b g(t-x) h(x) dx$$

where $g(t)$ is the Green's function for $\tilde{L}$ on D . Note that if we take $h(t) = \delta(t)$

$$y(t) = \int_a^b g(t-x) \delta(x) dx = g(t)$$

so the Green's function $g(t)$ can be defined as

$$\tilde{L}_g(t) = \delta(t)$$

Let us assume now that

$$f(x) = \int_a^b k(x, x') u(x') dx' \quad (114)$$

where $x = a$ is the position at the boundary along $-\omega$. Applying the previous assumptions to Eq. 113, we obtain

$$\alpha(x) k(x, x') + \frac{d}{dx} k(x, x') = \alpha(x) \delta(x - x') \quad (115)$$

for $x \neq x'$ the right side is zero and, therefore,

$$k(x \neq x', x') = c. \exp\left(-\int_{x'}^x \alpha(t) |dt|\right) \quad (116)$$

with c is a constant in the regions $x < x'$ and $x > x'$ where the solution is expected to be continuous. The δ function implies a step function in c for which the Green's function of the operator L_x is

$$G_w(x, \mathbf{r}') \begin{cases} 0 & \text{if } x < x' \\ \beta(\mathbf{r}') \exp\left(-\underbrace{\int_{x'}^x \beta(\mathbf{r}' + (t-x')\omega) dt}_{-\tau(\mathbf{r}, \mathbf{r}')} \right) & \text{otherwise} \end{cases} \quad (117)$$

with the optical density $\tau(\mathbf{r}, \mathbf{r}')$ between $\mathbf{r}$ and $\mathbf{r}'$. We obtain the convolution for by integrating $x \in [a, b]$ where b is the distance from $\mathbf{r}'$ to the boundary of the medium along the direction ω . Then,

$$G_\omega(\mathbf{r}, \mathbf{r}') = \beta(\mathbf{r}') \exp(-\tau(\mathbf{r}, \mathbf{r}')) \frac{\delta\left(\omega - \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\right)}{|\mathbf{r}-\mathbf{r}'|^2} \quad (118)$$

The δ function selects $\mathbf{r}$ that fits to the direction ω and the reference point $\mathbf{r}'$. Then, Eq. 112 can be written in terms of the collision density $f(\mathbf{r}, \omega) = \beta(\mathbf{r})I(\mathbf{r}, \omega)$ by the convolution with $G_\omega(\mathbf{r}, \mathbf{r}')$ and multiplication with $\beta(\mathbf{r})$:

$$f(\mathbf{r}, \omega) = \int_A \int_{4\pi} \frac{\Lambda(\mathbf{r}')}{4\pi} P(\mathbf{r}', \omega \cdot \omega') \times f(\mathbf{r}', \omega') \beta(\mathbf{r}) \exp(-\tau(\mathbf{r}, \mathbf{r}')) \times \frac{\delta\left(\omega - \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\right)}{|\mathbf{r}-\mathbf{r}'|^2} d\omega' d\mathbf{r}' + \Psi(\mathbf{r}, \omega) \quad (119)$$

where $\Psi(\mathbf{r}, \omega) = \Psi_{thermal}(\mathbf{r}, \omega) + \Psi_{solar}(\mathbf{r}, \omega)$

$$\Psi_{thermal}(\mathbf{r}, \omega) = \beta(\mathbf{r}) \int_A \beta_a(\mathbf{r}') B_\lambda(T(\mathbf{r}')) \times \exp(-\tau(\mathbf{r}, \mathbf{r}')) \frac{\delta\left(\omega - \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\right)}{|\mathbf{r}-\mathbf{r}'|^2} d\mathbf{r}' \quad (120)$$

and

$$\Psi_{solar}(\mathbf{r}, \omega) = F_0(\lambda) \beta(\mathbf{r}) \int_A \Theta_{sun}(\mathbf{r}') \times \exp(-\tau(\mathbf{r}, \mathbf{r}')) \frac{\delta\left(\omega - \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\right)}{|\mathbf{r}-\mathbf{r}'|^2} d\mathbf{r}' \quad (121)$$

Then the transformation of the radiative transfer equation (RTE) into an integral equation (IRTE) yields

$$f(\mathbf{r}, \omega) \int_A \int_{4\pi} k[(\mathbf{r}', \omega') \rightarrow (\mathbf{r}, \omega)] f(\mathbf{r}', \omega') d\omega' d\mathbf{r}' + \Psi(\mathbf{r}, \omega) \quad (122)$$

where the transport kernel k is

$$k[(\mathbf{r}', \omega') \rightarrow (\mathbf{r}, \omega)] = k_p[(\mathbf{r}', \omega') \rightarrow (\mathbf{r}, \omega)] \frac{\delta\left(\omega - \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\right)}{|\mathbf{r}-\mathbf{r}'|^2} \quad (123)$$

with the transport transition density $k_p[(\mathbf{r}', \omega') \rightarrow (\mathbf{r}, \omega)]$ given by

$$k_p[(\mathbf{r}', \omega') \rightarrow (\mathbf{r}, \omega)] = \frac{\Lambda(\mathbf{r}')}{4\pi} P(\mathbf{r}', \omega \cdot \omega') f(\mathbf{r}', \omega') \beta(\mathbf{r}) \exp(-\tau(\mathbf{r}, \mathbf{r}')) \quad (124)$$

which can be interpreted as a probability density of photons to perform the transition $(\mathbf{r}', \omega') \rightarrow (\mathbf{r}, \omega)$. The factor k and k_p can be found by integration over the medium.

Let us rewrite Eq. 122 in operator form as

$$f = Kf + \Psi \quad (125)$$

whose solution can be represented as a Neumann series (i.e., by the order of scattering):

$$f = \Psi + K\Psi + K^2\Psi + K^3\Psi + \dots = \sum_{n=0}^{\infty} K^n \Psi \quad (126)$$

which converges for all bounded media with $0 \leq \Lambda \leq 1$.

If the forward Monte Carlo for radiative transfer is based on IRTE, the backward Monte Carlo is based on the *principle of reciprocity* that describes the radiative process in the reverse direction. Applying the time reversal operator, that is, introducing the minus sign before the direction ω in Eq. 112, we obtain the so-called *adjoint RTE* containing the adjoint transport operator K^+ . Thus, the solution for an adjoint integrodifferential equation with delta function source and homogeneous boundary condition is

$$f^+ = K^+ f^+ + \Psi \quad (127)$$

where Ψ is the detector function. The solution of Eq. 127 can be found as for forward time direction:

$$f^+ = \sum_{\eta=0}^{\infty} K^{+\eta} \cdot \Psi \quad (128)$$

showing the solution of the RTE in the form of Newman series can be expressed by nested integrals.

Monte Carlo methods are ideal for computing numerical solutions of previous integrals. Considering a natural event, the set of possible results can be continuous or discrete. The probability of an event can be measured using the so-called probability density function *pdf*, while the cumulative distribution function *cdf* is the integral of *pdf*. All the terms we have introduced, that is, the phase function and the free path length, can be computed using the *pdf* and *cdf* from which it is also possible to derive the expectation value, higher momentum, variance, and standard deviation of a random variable. The physics of radiation transport is merged with stochastic methods to create

numerical algorithms that estimate the collision density of the radiance. The methods described are based on the principle of reciprocity, either to simulate the natural process in forward direction, the photon source in the sun, or to describe the backward or adjoint process, the trajectory originating at the detector's position. In this case, the initial distribution of propagation corresponds to the field of view (FOV) of the detector. One key for the implementation of Monte Carlo simulation is the ray tracer simulating the propagation of light. A ray tracer uses random numbers to decide over the occurrence of events like extinction of light, scattering, and absorption, according to the local physical conditions. The trajectory is completed when the photon is absorbed or leaves the atmosphere.

Monte Carlo methods are exhaustively described in books by Marchuck et al. (1980) and Marshak and Davis (2005). A series of different papers deals with different aspects of Monte Carlo methods; among the others we cite Iwabuchi (2006) and Baker et al. (2003), which describe numerical techniques to reduce the computational demands of Monte Carlo simulations; O'Hirok and Gautier (1998), which describes techniques for broadband Monte Carlo simulations; and Cornet et al. (2010) and Battaglia and Mantovani (2005), which describes techniques to include polarization into Monte Carlo simulations. The latter also focuses on microwave radiation and anisotropic media. Tanaka and Ellison (1996; 2000) describe thermal infrared Monte Carlo simulations.

Summary

The radiative transfer equations in an atmosphere-surface coupled system are a complex integrodifferential equations system requiring proper solution techniques. Despite advances in computer science, solutions remain time consuming and, to ensure the required accuracy, they require proper selection. This entry deals with some possible solution techniques. In the Introduction, we cite the classical texts and have selected the most advanced one in order to give the most effective overview. Further, we describe the selected methods by a cascade process by which the different techniques can be used singly or all together. The methods we have selected are the Fourier expansion, the iterative method, primary scattering, Gauss-Seidel, Discrete Ordinate, and Monte Carlo. All these methods are interconnected; the solution of one method can be also used for another method, as will be explained in the text. Even the statistical approach or Monte Carlo uses some solutions described by the other numerical selected methods. We introduce the rigorous analytical approach by the solutions embedded into the iterative and Fourier approach. In order to give a complete picture, we also introduce the input models of atmosphere and surface and how to use them. The bibliography provides sources for some of the software, which is usually in the public domain.

Bibliography

- Barker, H. W., Goldstein, R. K., and Stevens, D. E., 2003. Monte Carlo simulation of solar reflectances for cloudy atmospheres. *Journal of the Atmospheric Sciences*, **60**(16), 1881–1894.
- Battaglia, A., and Mantovani, S., 2005. Forward Monte Carlo computations of fully polarized microwave radiation in non isotropic media. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **95**, 285–308.
- Berk, A., Bernstein, L. S., and Robertson, D. C., 1989. *Modtran MODTRAN: A Moderate Resolution Model for LOWTRAN 7*. Burlington: Spectral Sciences.
- Bohren, C. F., and Clotiaux, E. E., 2006. *Fundamentals of Atmospheric Radiation*. Weinheim: Wiley.
- Chandrasekhar, S., 1960. *Radiative Transfer*. New York: Dover.
- Cornet, C., C-Labonnote, L., and Szczap, F., 2010. Three-dimensional polarized Monte Carlo atmospheric radiative transfer model (3DMCPOL): 3D effects on polarized visible reflectances of a cirrus cloud. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **111**, 174–186, doi:10.1016/j.jqsrt.2009.06.013.
- De Haan, J. F., Bosma, P. B., and Hovenier, J. W., 1987. The adding method for multiple scattering calculations of polarized light. *Astronomy and Astrophysics*, **183**, 371–391.
- Goody, R. M., and Yung, Y. L., 1989. *Atmospheric Radiation Theoretical Basis*. New York: Oxford University Press.
- <http://code.google.com/p/i3rc-monte-carlo-model/>
- IAMAP Radiation Commission, 1975. *Standard Procedures to Compute Atmospheric Radiative Transfer in a Scattering Atmosphere*. Boulder: NCAR, Vol. I.
- IAMAP Radiation Commission, 1980. *Standard Procedures to Compute Atmospheric Radiative Transfer in a Scattering Atmosphere*. Boulder: NCAR, Vol. II.
- Iwabuchi, H., 2006. Efficient Monte Carlo methods for radiative transfer modeling. *Journal of the Atmospheric Sciences*, **63**(9), 2324–2339.
- Kneizys, F. X., Shettle, E. P., Abreu, L. W., Chetwynd, J. H., and Anderson Lowtran, G. P., 1988. *Users Guide to LOWTRAN 7*. Hanscom AFB: Air Force Geophysics Lab.
- Kondratiev, K. Y., Kozoderov, V. V., and Smokty, O. I., 1992. *Remote Sensing of the Earth from Space: Atmospheric Correction*. Heidelberg: Springer.
- Lenoble, J., 1985. *Radiative Transfer in Scattering and Absorbing Atmospheres: Standard Computational Procedures*. Hampton: Deepak Publishing, p. 300.
- Lenoble, J., 1993. *Atmospheric Radiative Transfer*. Hampton: Deepak Publishing, p. 532.
- Levoni, C., Cervino, M., Guzzi, R., and Torricella, F., 1997. Atmospheric aerosol optical properties: a data base of radiative characteristics for different component and classes. *Applied Optics*, **36**, 8031–8041.
- Levoni, C., Cattani, E., Cervino, M., Guzzi, R., and Di Nicolantonio, W., 2001. Effectiveness of the MS-method for computation of the intensity field reflected by a multilayer plane-parallel atmosphere: results from an accelerated yet accurate radiative transfer code. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **4**, 636–649.
- Liou, K. N., 1980. *An Introduction to Atmospheric Radiation*. New York: Academic.
- Liou, K. N., 1992. *Radiation and Cloud Processes in the Atmosphere*. New York: Oxford University Press.
- Marchuk, G., Mikhailov, G., Nazaraliev, M., Darbinjan, R., Kargin, B., and Elepov, B., 1980. *The Monte Carlo Methods in Atmospheric Optics*. Berlin: Springer, p. 208.
- Marshak, A., and Davis, A. B. (eds.), 2005. *3D Radiative Transfer in Cloudy Atmospheres*. Springer: Berlin, p. 686.
- Mayer, B., Emde, C., Buras, R., Hamann, U., and Kylling, A., 2005. LibRadTran: library for radiative transfer. <http://www.libradtran.org/doku.php>.
- Miskolczi, F., Rizzi, R., Guzzi, R., and Bonzagni, M., 1998. A new high resolution transmittance code and its application in the field of remote sensing. In Lenoble, J., and Geleyn, J. F., (eds.), *International Radiation Symposium*. Deepak Publishing: Hampton VA, USA.
- Nakajima, T., and Tanaka, M., 1981. Algorithms for radiative intensity calculations in moderate thick atmospheres using a truncation approximation. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **40**, 51–69.
- O'Hirok, W., and Gautier, C., 1998. A three-dimensional radiative transfer model to investigate the solar radiation within a cloudy atmosphere. Part I: spatial effects. *Journal of the Atmospheric Sciences*, **55**(12), 2162–2179.
- Petty, G. W., 2006. *A First Course in Atmospheric Radiation*, 2nd edn. Madison: Sundog.
- Pincus, R., and Evans, K. F., 2009. Computational cost and accuracy in calculating three-dimensional radiative transfer: results for new implementations of Monte Carlo and SHDOM. *Journal of the Atmospheric Sciences*, **66**(10), 3131–3314.
- Preisendorfer, R., 1965. *Radiative Transfer on Discrete Spaces*. New York: Pergamon.
- Ricchiazzi, P., Shiren, Y., and Gautier, C., 2007. SBDART: a practical tool for plane-parallel radiative transfer in the earth's atmosphere. Earth Space Research Group, Institute for Computational Earth System Science University of California, Santa Barbara, <http://www.paulschou.com/tools/sbdart/>.
- Sobolev, V., 1975. *Light Scattering in Planetary Atmospheres*. New York: Pergamon. 442p.
- Spurr, R. J., 2001. *A General Discrete Ordinate Approach to the Calculation of Radiances and Analytical Weighting Functions with Applications to Atmospheric Remote Sensing*. PhD thesis, Eindhoven Tech. University.
- Stamnes, K., 1982. On the computation of angular distributions of radiation in planetary atmosphere. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **28**, 47–51.
- Stamnes, K., and Swanson, R., 1981. A new look at the discrete ordinate method for radiative transfer calculation in anisotropically scattering atmosphere. *Journal of the Atmospheric Sciences*, **38**, 387–399.
- Stamnes, K., Tsay, S. C., Wiscombe, W. J., and Jayaweera, K., 1988. Numerical stable algorithm for discrete ordinate radiative transfer in multiple scattering and emitting layered media. *Applied Optics*, **27**, 2502–2509.
- Takara, E. E., and Ellingson, R. G., 1996. Scattering effects on longwave fluxes in broken cloud fields. *Journal of the Atmospheric Sciences*, **53**(10), 1464–1476.
- Takara, E. E., and Ellingson, R. G., 2000. Broken cloud field longwave-scattering effects. *Journal of the Atmospheric Sciences*, **57**(9), 1298–1310.
- Thomas, G. E., and Stamnes, K., 1999. *Radiative Transfer in Atmosphere and Ocean*. Cambridge: Cambridge University Press, p. 517.
- Van De Hulst, H. C., 1980. *Multiple Light Scattering. Tables, Formulas and Applications*. New York: Academic.
- Verstraete, M. M., Pinty, B., and Dickinson, R. E., 1990. A physical model of the bidirectional reflectance of vegetation canopies. I – Theory II – inversion and validation. *Journal of Geophysical Research*, **95**, 11755–11775. (ISSN 0148–0227), Research supported by ESA, CNRS, and NCAR.
- World Climate Programme. 1982. WCP-43 Tropospheric aerosols: review and current data on physical and optical properties (computed by Harris, F. S., and Gerber, H. G.), WMO Geneve.
- Yanovithskij, E. G., 1997. *Light Scattering in Inhomogeneous Atmosphere*. Heidelberg: Springer.

Cross-references

[Aerosols](#)
[Cloud Properties](#)
[Fields and Radiation](#)
[Land-Atmosphere Interactions, Evapotranspiration](#)
[Land Surface Roughness](#)
[Ocean-Atmosphere Water Flux and Evaporation](#)
[Optical/Infrared, Scattering by Aerosols and Hydrometeors](#)
[Radiative Transfer, Theory](#)

RADIATIVE TRANSFER, THEORY

Frank S. Marzano

Department of Information Engineering, Sapienza
 University of Rome, Rome, Italy
 Centre of Excellence CETEMPS, University of L'Aquila,
 L'Aquila, Italy

Definition

The radiative transfer theory, also called transport theory, is the theory describing the wave propagation through a medium characterized by a random distribution of scatterers. It usually applies to electromagnetic radiation, but it can be generalized to acoustic radiation. The radiative transfer theory is one of the two theories developed to deal with radiation absorption and multiple-scattering problems: the other approach is the so-called wave analytical theory. It is generally formulated in terms of specific intensity and can be extended to polarized radiation. The radiative transfer theory constitutes the physical basis of several remote sensing techniques due to its relative simplicity and capability to deal with multiple-scattering effects.

Introduction

The subject of radiative transfer is very transversal and, indeed, covers several research fields, including astrophysics, applied physics, optics, planetary sciences, atmospheric sciences, meteorology, and engineering. Generally speaking, we can define the *radiative transfer theory* (RTT), also called *transport theory*, as the theory describing the wave propagation through a medium characterized by a *random distribution of scatterers* (Chandrasekhar, 1960). It usually applies to electromagnetic radiation, but it can be generalized to acoustic radiation (Ishimaru, 1978; Tsang et al., 2000).

The RTT is one of the two distinct theories developed to deal with radiation absorption and multiple-scattering problems: the other approach is the so-called wave analytical theory (WAT). The latter starts, in case of electromagnetic waves, from the basic *Maxwell differential equations* which are usually transformed into an equivalent wave equation (Ishimaru, 1978; Tsang et al., 1985). The wave analytical theory is, in principle, able to treat absorption and scattering to any order together with diffraction and interference effects. However, as intuitive, the detailed approach of the WAT is very cumbersome and mathematically complicated,

so that a formulation which includes all multiple-scattering terms is practically impossible to derive.

The RTT, conversely, does not start with the wave equation, but simplifies the problem dealing with the *transport of energy* through a medium containing scatterers (Chandrasekhar, 1960; Ishimaru, 1978; Ulaby et al., 1986). The development of theory is almost heuristic as it is not rigorously derived from the wave equation as the wave analytical theory. Although diffraction and interference effects are included in the description of the absorption and scattering properties of a single scatterer, RTT does not actually include the *interferential effects* of the ensemble of the scatterers. As a matter of fact, RTT assumes that there is no correlation between the fields scattered by each particle (so-called independent-scattering assumption), so that the addition of the scattered power, rather than the addition of the scattered fields, holds. Indeed, RTT carries information on the mutual coherence of the diffuse field, as it will be introduced later on dealing with the assumptions and limitations of RTT at the end of this section (Ishimaru, 1975; Furutsu, 1975).

The RTT can be expressed in terms of specific intensity which expresses the radiant power per unit area, solid angle, and frequency. If the specific intensity is scalar, as for the case of totally unpolarized light, the *radiative transfer equation* (RTE) becomes a single differential equation which is usually called the *scalar RTE* (Ulaby and Elachi, 1990). For vector fields, the scalar specific intensity can be generalized using the four Stokes parameters which lead to four coupled differential equations. The latter is sometimes called the *vector RTE* (Tsang et al., 1985). In the following sections, we will briefly introduce both scalar and vector RTEs.

Brief history of radiative transfer theory

The RTT was pioneered by astrophysicists who opened the field at the beginning of the twentieth century. The work of Schuster (1905) on the investigation of the light radiation transfer through a foggy atmosphere appears to be the first paper discussing the importance of multiple scattering. Schuster basically formulated the problem in terms of upward and downward light beams, the basic assumption for the so-called two-stream approximation of radiative transfer. The latter was also employed by Schwarzschild (1906) to explain the limb darkening of the sun. In his effort to understand the physical interior of a star, Eddington (1916) developed a first-order expansion of intensity in terms of Legendre polynomials, leading to the so-called Eddington approximation of radiative transfer. In the same years, Schwarzschild introduced the concept of the medium emission as well as absorption within the context of thermodynamic equilibrium, a subject widely used to deal with thermal infrared radiation in molecular atmosphere with negligible scattering. These pioneering works were collected in a volume by Menzel in the middle of the 1960s (1966).

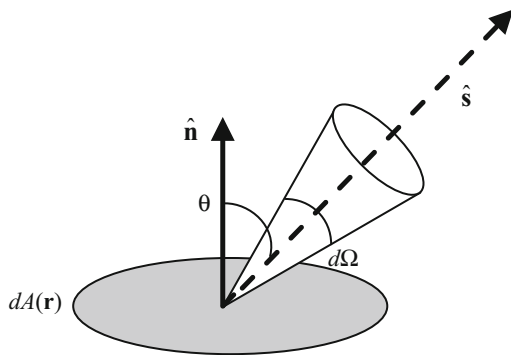
In the 1950s, in his landmark book, Chandrasekhar presented the subject of radiative transfer in plane-parallel atmospheres as a branch of mathematical physics and developed numerous solution methods and techniques, including the consideration of polarized radiation (Chandrasekhar, 1960). An equally fundamental treatise is the one of Sobolev (1975). In the 1960s and 1970s, many efforts were carried out to provide a robust physical basis of the RTT and its relation with the wave analytical theory (Ishimaru, 1978; Tsang et al., 1985).

The basic differential equation of RTT is equivalent to Boltzmann's equation (also known as the Boltzmann-Maxwell collision equation), used in the kinetic theory of gases (Sommerfeld, 1956). There is a noteworthy equivalence between RTT and the neutron transport theory which was well established by Davison (1958). The formulation of RTT is, indeed, flexible and capable of treating many physical problems. It has been successfully applied to the problems of atmospheric radiation, underwater visibility, marine biology, photographic optics, and wave propagation through planets, stars, and galaxies.

Specific intensity

The fundamental quantity in the RTT is the *scalar specific intensity* I , sometimes also called radiance (indicated by L) or brightness (indicated by B). The concept of the specific intensity can be introduced considering a flow of wave energy in a random medium where the frequency, phase, and amplitude of the wave undergo random variations in time and space. This implies that the magnitude and direction of its power flow density vector (also called Poynting vector in electromagnetics) vary continuously in time and space (Ishimaru, 1978; Tsang et al., 1985).

At a given point $\mathbf{r}$ and a direction defined by the unit vector $\hat{\mathbf{s}}$, the scalar specific intensity $I(\mathbf{r}, \hat{\mathbf{s}})$ is measured in $\text{W m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}$ (where sr stands for steradian or solid unit angle). Referring to Figure 1, the amount of elementary power dP (in Watts) flowing within a solid angle $d\Omega$ through an elementary area dA oriented along



Radiative Transfer, Theory, Figure 1 Geometry for a specific intensity emitted by an elementary surface dA .

the direction of the unit vector $\hat{\mathbf{n}}$ in a frequency interval dv is given by

$$dP = I(\mathbf{r}, \hat{\mathbf{s}}) \cos \theta dA d\Omega dv \quad (1)$$

Indeed, Equation 1 describes the radiation flux emitted from a surface dA along $\hat{\mathbf{s}}$ and it is often referred as the *forward* (or *outward*) specific intensity $I^+ = I(\mathbf{r}, \hat{\mathbf{s}})$ (also called *surface intensity*). Conversely, if the point $\mathbf{r}$ is taken on the surface dA , we can express the *backward* (or *inward*) specific intensity $I^- = I(\mathbf{r}, -\hat{\mathbf{s}})$ incident upon the surface along $-\hat{\mathbf{s}}$ (also called *field intensity*). In space, the inward power upon dA should be identical to the outward power emitted from the same surface, so that: $I^+ = I^-$.

In electromagnetics, the specific intensity $I(\mathbf{r}, \hat{\mathbf{s}})$ can be statistically related to the Poynting vector $\mathbf{S}$. If the latter at a given point $\mathbf{r}$ is a random function of time and each component of $\mathbf{S}$ varies in time, the tip vector of $\mathbf{S}$ moves randomly in time. If this random variability is expressed by a probability density function p_S , we can then define the specific intensity as the *ensemble average value* of the random vector $\mathbf{S}$. This means that the $I(\mathbf{r}, \hat{\mathbf{s}})$ is basically the sum of all random Poynting vectors $\mathbf{S}$ whose tips are located within a solid unit angle in the direction $\hat{\mathbf{s}}$. The same interpretation can be easily extended to acoustic waves (Ishimaru, 1978).

Flux and energy density

The *forward flux density* F^+ (measured in $\text{W m}^{-2} \text{Hz}^{-1}$) passing through an elementary area dA on a finite surface A is obtained by integrating Equation 1 over a semisolid angle $2\pi^+$ in the forward interval ($0 \leq \theta \leq \pi/2$) along the direction $\hat{\mathbf{n}}$:

$$F^+(\mathbf{r}, \hat{\mathbf{n}}) = \int_{2\pi^+} I(\mathbf{r}, \hat{\mathbf{s}}) \hat{\mathbf{s}} \cdot \hat{\mathbf{n}} d\Omega \quad (2)$$

Similarly, for the *backward flux density* F^- through dA in the backward direction ($-\hat{\mathbf{n}}$) and for a semisolid angle $2\pi^-$ in the backward interval ($\pi/2 \leq \theta \leq \pi$), we can define:

$$F^-(\mathbf{r}, \hat{\mathbf{n}}) = \int_{2\pi^-} I(\mathbf{r}, \hat{\mathbf{s}}) \hat{\mathbf{s}} \cdot (-\hat{\mathbf{n}}) d\Omega \quad (3)$$

For a radiating surface, the forward flux density F^+ is often called *radiant emittance* or *radiant excitance* (Elachi, 1987). For receiving surface, the backward flux density F^- is also called *irradiance*. The net flux density can be expressed as the component of a flux density vector $\mathbf{F}(\mathbf{r})$ along $\hat{\mathbf{n}}$ equal to the difference between the forward flux density and the backward flux density. For the flux density vector $\mathbf{F}(\mathbf{r})$, it holds over the whole solid angle:

$$F(r) = \int_{4\pi} I(r, \hat{\mathbf{s}}) \hat{\mathbf{s}} d\Omega \quad (4)$$

Other useful definitions related to the specific intensity are the *energy density* $u(\mathbf{r})$ and the *average intensity* $U(\mathbf{r})$, defined by

$$u(\mathbf{r}) = \frac{1}{c} \int_{4\pi} I(\mathbf{r}, \hat{\mathbf{s}}) d\Omega = \frac{c}{4\pi} U(\mathbf{r}) \quad (5)$$

where c is the wave velocity with u measured in $\text{J m}^{-3} \text{Hz}^{-1}$ and U measured in $\text{W m}^{-2} \text{Hz}^{-1}$. Note that Equation 4 is derived considering that the energy density $du(\mathbf{r})$ can be expressed as the ratio of the energy $I dA d\Omega dv dt$ and the volume $dA c dt$ it should occupy.

As an important example, we can consider the case when the specific intensity $I(\mathbf{r}, \hat{\mathbf{s}})$ is independent of the direction $\hat{\mathbf{s}}$ so that the radiation is said to be *isotropic*. In this circumstance, the total net flux density F along $\hat{\mathbf{n}}$ can be expressed by

$$F(\mathbf{r}) = \mathbf{F}(\mathbf{r}) \cdot \hat{\mathbf{n}} = \int_{4\pi} I(\mathbf{r}, \hat{\mathbf{s}}) \cos \theta d\Omega = I_0(\mathbf{r}) \pi \quad (6)$$

where I_0 is average specific intensity value. The previous relationship is called the *Lambert law* which can be also expressed in terms of power from Equation 1 as a cosine law: $P = P_0 \cos \theta$ with P_0 the average power.

Invariance and boundary conditions

It can be proved that the specific intensity is *invariant* along the ray path in free space (Ishimaru, 1978; Tsang et al., 2000). Even though this property might seem to contradict the intuitive spreading of power flux in free space, it should be reminded that the specific intensity is defined per unit solid angle from Equation 1. The proof of the *specific intensity invariance* can be approached by considering two points $\mathbf{r}_1$ and $\mathbf{r}_2$ in free space separated by a distance r along the direction $\hat{\mathbf{s}}$ and two small areas dA_1 and dA_2 perpendicular to $\hat{\mathbf{s}}$. Let $I(\mathbf{r}_1, \hat{\mathbf{s}})$ and $I(\mathbf{r}_2, \hat{\mathbf{s}})$ be the specific intensity at $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. The power received by dA_2 can be expressed in two ways: (1) as $I(\mathbf{r}_1, \hat{\mathbf{s}}) dA_1 d\Omega_1 dv$ from Equation 1 and (2) as $I(\mathbf{r}_2, \hat{\mathbf{s}}) dA_2 d\Omega_2 dv$. The previous expression must be equal, so that $I(\mathbf{r}_1, \hat{\mathbf{s}}) dA_1 d\Omega_1 = I(\mathbf{r}_2, \hat{\mathbf{s}}) dA_2 d\Omega_2$. But it holds $dA_1 = r^2 d\Omega_2$ and $dA_2 = r^2 d\Omega_1$, so that it results $I(\mathbf{r}_1, \hat{\mathbf{s}}) = I(\mathbf{r}_2, \hat{\mathbf{s}})$.

Once dealing with applications, the random medium is very likely inhomogeneous. This means that when looking for the specific intensity within the considered volume, we need to know the *boundary surface conditions* in terms of specific intensity (Tsang et al., 1985; Tsang et al., 2000; Ulaby et al., 1986). It is customary in RTT to divide the boundary conditions into three categories: (1) isotropic surface, (2) planar surface, and (3) rough surface.

In the first case, the surface is also called *Lambertian* as the behavior of the specific intensity $I(\mathbf{r}, \hat{\mathbf{s}})$ is independent of the direction $\hat{\mathbf{s}}$ and follows that of the Lambert law. This can be expressed in terms

$$I^+(\mathbf{r}_s) = I^-(\mathbf{r}_s) \quad (7)$$

where $\mathbf{r}_s$ is the position vector belonging to the boundary surface S .

In the second case, the surface is called *Fresnelian* as the behavior of the specific intensity $I(\mathbf{r}, \hat{\mathbf{s}})$ is regulated by the Snell law (i.e., reflection angle equal to incident angle) and the relation between the reflected and transmitted wave amplitude, with respect to the incident one, follows the Fresnel coefficient (see Figure 2). If $\Gamma_{\parallel}$ and $\Gamma_{\perp}$ are the reflection coefficients of the electric field polarized parallel and perpendicular to the plane of incidence, respectively, for unpolarized (scalar) incident specific intensity I^- and reflected specific intensity I^+ , it holds

$$I^+(\mathbf{r}_s, \hat{\mathbf{s}}_r) = |\Gamma(\hat{\mathbf{s}}_i)|^2 I^-(\mathbf{r}_s, \hat{\mathbf{s}}_i) \text{ with } \hat{\mathbf{n}} \cdot \hat{\mathbf{s}}_r = \hat{\mathbf{n}} \cdot \hat{\mathbf{s}}_i \quad (8)$$

where $\hat{\mathbf{s}}_r$ and $\hat{\mathbf{s}}_i$ are the unit vector for the reflection and incident directions, whereas $|\Gamma|^2 = 0.5 (|\Gamma_{\parallel}|^2 + |\Gamma_{\perp}|^2)$ is the unpolarized reflectance (or reflectivity) of the Fresnel boundary surface (Tsang et al., 1985). For the *surface transmittance*, a similar argumentation can be carried out considering the transmission Fresnel coefficients and the power conservation at the surface so that the incident power should be equal to the sum of the reflected and transmitted ones (Ishimaru, 1978).

In the third general case, the surface may be called *natural* and there is not necessarily an analytical law relating the incident and the reflected specific intensity (Ulaby et al., 1986). If $\gamma_d(\hat{\mathbf{s}}_r, \hat{\mathbf{s}}_i)$ is the power differential reflectivity of the boundary surface, it can be written in a general way:

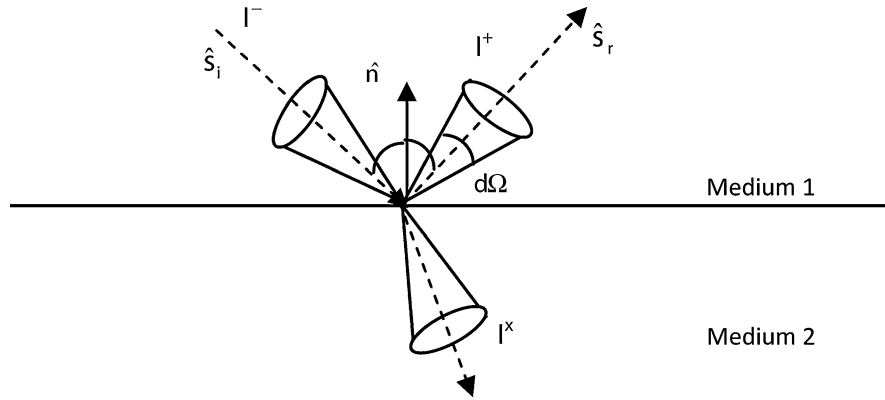
$$I^+(\hat{\mathbf{r}}_s, \hat{\mathbf{s}}_r) = \int_{2\pi} \gamma_d(\hat{\mathbf{s}}_r, \hat{\mathbf{s}}_i) I^-(\mathbf{r}_s, \hat{\mathbf{s}}_i) d\Omega_i \quad (9)$$

where $d\Omega_i$ is the solid angle including all incident directions in 2π ($\pi/2 \leq \theta \leq \pi$) and $\hat{\mathbf{s}}_r$ spans the forward interval ($0 \leq \theta \leq \pi/2$). The characterization of γ_d will depend on the surface geometry and composition.

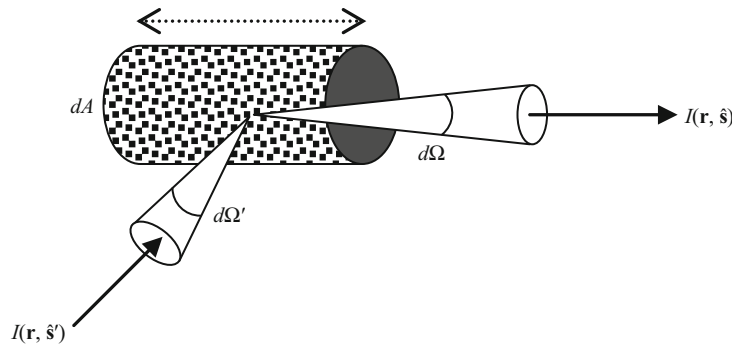
Radiative transfer equation

After examining the invariance property and the boundary conditions, we are ready to analyze the propagation of the fundamental quantity in RTT, the specific intensity $I(\mathbf{r}, \hat{\mathbf{s}}_r)$, through a medium containing a random distribution of scatterers. The latter scatter, absorb, and emit the wave energy, and all these characteristics should be included into a differential equation expressed in terms of the specific intensity.

The scalar *radiative transfer equation* (RTE) can be derived by considering an unpolarized specific intensity $I(\mathbf{r}, \hat{\mathbf{s}})$ incident upon a cylindrical elementary volume dV with a cross section dA and a length ds , as in Figure 3. The volume $dV = ds dA$ contains a random distribution of scatterers with different properties described by the parameters q_i ($i = 1 - Q$). The latter can represent the size, shape, and orientation variability of the scatterers and are



Radiative Transfer, Theory, Figure 2 Boundary condition for specific intensity at the planar surface S separating two media with different refractive indexes.



Radiative Transfer, Theory, Figure 3 Propagation and interaction of the specific intensity upon an elementary volume containing a random distribution of scatterers.

usually described by a *joint* probability density function $p_q(q_1, \dots, q_Q)$, usually normalized to one. In case we refer to spherical particles with different diameters, then the Q scatterer parameters reduce to $q_1 = D$ and $p_D(D)$. In order to derive the RTE, we need (1) first, to characterize the volumetric interaction in terms of absorption and scattering and (2) then to consider the balance of the power along the ray, taking into account thermal emission, incident radiation, and multiple scattering.

Volumetric interaction parameterization

The volumetric interaction between the incident radiation and an ensemble of scatterers needs to take into account both absorption and scattering. In this respect, we can exploit the assumption of independent scattering, which is a negligible correlation among the scattered fields. This means that the absorbed and scattered power of each scatterer can be simply summed up to provide the total absorbed and scattered power (Ishimaru, 1978; Tsang et al., 2000).

For a single scatterer at a point $\mathbf{r}$ and characterized by some properties q_i , we can define the absorption cross section $\sigma_a(r, \hat{s}_i, q_1, \dots, q_Q)$ (measured in m^2) with respect to

an incident wave along the unit direction $\hat{s}_i$. The cross section σ_a is the ratio between the power dP_a , absorbed by the scatterer itself, and incident power flux density S_i and will depend on the scatterer geometry, orientation, and composition. Considering an ensemble of the previous scatterers, randomly distributed after $p_q(q_1, \dots, q_Q)$, and summing over all the absorbed powers dP_a , the *volumetric absorption coefficient* $\kappa_a(\mathbf{r}, \hat{s}_i)$ is defined as

$$\kappa_a(\mathbf{r}, \hat{s}_i) = N_q \int_{\Delta q} \sigma_a(\mathbf{r}, \hat{s}_i, q_1, \dots, q_Q) p_q(q_1, \dots, q_Q) dq_1, \dots, dq_Q \quad (10)$$

where N_q is the number of scatterers per unit volume (measured in m^{-3}) and Δq is the domain of integration. This means that κ_a is measured in m^{-1} .

A similar argument can be used to introduce the *differential scattering cross section* $\sigma_d(\mathbf{r}, \hat{s}_r, \hat{s}_i, q_1, \dots, q_Q)$ (measured in m^2) which, however, depends not only on the incident direction $\hat{s}_i$ but also on the scattering direction $\hat{s}_r$. The cross section σ_d is given by the ratio of the power flux intensity $r^2 dS_r$, scattered at a distance r in the far field by the scatterer itself, and the incident power flux density S_i

along $\hat{\mathbf{s}}_i$. Considering a scatterer ensemble and summing over all the scattered power flux intensities $r^2 dS_r$, the volumetric *differential scattering coefficient* $\kappa_d(\mathbf{r}, \hat{\mathbf{s}}_r, \hat{\mathbf{s}}_i)$ may be defined as

$$\kappa_d(\mathbf{r}, \hat{\mathbf{s}}_r, \hat{\mathbf{s}}_i) = N_q \int_{\Delta q} \sigma_d(\mathbf{r}, \hat{\mathbf{s}}_i, q_1, \dots, q_Q) \quad (11)$$

$$p_q(q_1, \dots, q_Q) dq_1 \dots dq_Q$$

From Equation 11, it is straightforward to derive the *volumetric scattering coefficient* $\kappa_s(\mathbf{r}, \hat{\mathbf{s}}_i)$, as it is the integral of κ_d over all scattering direction in 4π :

$$\begin{aligned} \kappa_s(\mathbf{r}, \hat{\mathbf{s}}_i) &= \int_{4\pi} \kappa_d(\mathbf{r}, \hat{\mathbf{s}}_r, \hat{\mathbf{s}}_i) d\Omega_r = N_q \\ &\int_q \sigma_s(\mathbf{r}, \hat{\mathbf{s}}_i, q_1, \dots, q_Q) p_q(q_1, \dots, q_Q) dq_1 \dots dq_Q \end{aligned} \quad (12)$$

where Ω_r is the scattering solid angle including all directions $\hat{\mathbf{s}}_r$ and σ_s is the *scattering cross section* of each scattered, measured in m^2 . From Equations 11 and 12, it is also derived that the scattering cross section σ_s is the integral of differential cross section σ_d over the whole solid angle. Note that the *coefficient* $\kappa_s(\mathbf{r}, \hat{\mathbf{s}}_i)$, measured in m^{-1} , is consistent with the *coefficient* $\kappa_a(\mathbf{r}, \hat{\mathbf{s}}_i)$, whereas σ_s is the ratio between the scattered power dP_r and the incident power flux density S_i .

Finally, for an ensemble of scatterers at a point $\mathbf{r}$, where both volumetric absorption and scattering take place, it is customary to define the *volumetric extinction coefficient* $\kappa_e(\mathbf{r}, \hat{\mathbf{s}}_i)$ as

$$\kappa_e(\mathbf{r}, \hat{\mathbf{s}}_i) = \kappa_a(\mathbf{r}, \hat{\mathbf{s}}_i) + \kappa_s(\mathbf{r}, \hat{\mathbf{s}}_i) \quad (13)$$

where the extinction is supposed to be sum of the absorption and scattering effects. A further important quantity is the *volumetric scattering phase function* $p(\mathbf{r}, \hat{\mathbf{s}}_r, \hat{\mathbf{s}}_i)$, defined as

$$p(\mathbf{r}, \hat{\mathbf{s}}_r, \hat{\mathbf{s}}_i) = 4\pi \frac{\kappa_d(\mathbf{r}, \hat{\mathbf{s}}_r, \hat{\mathbf{s}}_i)}{\kappa_s(\mathbf{r}, \hat{\mathbf{s}}_i)} \quad (14)$$

where its angular normalization over the whole solid angle is 4π . Sometimes, the phase function is defined with respect to κ_e , instead of κ_s , so that its angular normalization yields the product between 4π and the *volumetric albedo* (e.g., Ishimaru, 1978). The latter is defined as

$$w(\mathbf{r}, \hat{\mathbf{s}}_i) = \frac{\kappa_s(\mathbf{r}, \hat{\mathbf{s}}_i)}{\kappa_e(\mathbf{r}, \hat{\mathbf{s}}_i)} \quad (15)$$

As an example, we can consider a random distribution of spherical particles of diameter D_0 . In this case, the joint probability distribution reduces to a Dirac impulse function: $p_D(D) = \delta(D - D_0)$. From Equations 10, 11, 12, 13, 14, and 15, the volumetric interaction parameters are then simply equal to

$$\begin{aligned} \kappa_a(\mathbf{r}, \hat{\mathbf{s}}_i) &= N_D \sigma_a(\mathbf{r}, \hat{\mathbf{s}}_i); \quad \kappa_s(\mathbf{r}, \hat{\mathbf{s}}_i) = N_D \sigma_s(\mathbf{r}, \hat{\mathbf{s}}_i); \\ w(\mathbf{r}, \hat{\mathbf{s}}_i) &= \frac{\sigma_s(\mathbf{r}, \hat{\mathbf{s}}_i)}{\sigma_e(\mathbf{r}, \hat{\mathbf{s}}_i)} \end{aligned} \quad (16)$$

Differential equation formulation

Considering the geometry of Figure 3, we can then suppose the random scatterers characterized by the volumetric interaction parameters previously defined. From Equation 13, the reduction of the specific intensity $I(\mathbf{r}, \hat{\mathbf{s}})$ due to its propagation through the volume $dV = ds dA$ is equal to

$$\begin{aligned} dI(\mathbf{r}, \hat{\mathbf{s}}) &= -ds[\kappa_a(\mathbf{r}, \hat{\mathbf{s}}) + \kappa_s(\mathbf{r}, \hat{\mathbf{s}})] \\ I(\mathbf{r}, \hat{\mathbf{s}}) &= -ds\kappa_e(\mathbf{r}, \hat{\mathbf{s}})I(\mathbf{r}, \hat{\mathbf{s}}) \end{aligned} \quad (17)$$

On the other hand, the specific intensity increases because a portion of $I(\mathbf{r}, \hat{\mathbf{s}}')$, incident on dV from another direction $\hat{\mathbf{s}}'$, is scattered into the direction $\hat{\mathbf{s}}$ and is added to $I(\mathbf{r}, \hat{\mathbf{s}})$. This contribution along $\hat{\mathbf{s}}$ at a point $\mathbf{r}$ can be derived from the *differential scattering coefficient* $\kappa_d(\mathbf{r}, \hat{\mathbf{s}}_r, \hat{\mathbf{s}}_i)$ by integrating $I(\mathbf{r}, \hat{\mathbf{s}}')$, weighted by $\kappa_d(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}')$, over all the incident directions $\hat{\mathbf{s}}'_i = \hat{\mathbf{s}}'$ within the entire solid angle:

$$dI(\mathbf{r}, \hat{\mathbf{s}}) = +ds \int_{4\pi} \kappa_d(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}') I(\mathbf{r}, \hat{\mathbf{s}}') d\Omega' \quad (18)$$

where Ω' is the solid angle including all directions incident on the volume dV from directions different from $\hat{\mathbf{s}}$. The term in Equation 18 is often called the *multiple-scattering pseudo-source*.

Within the volume dV , if a local thermodynamic equilibrium can be supposed, the *Kirchhoff law* holds. The latter states that a medium that absorbs radiation at a given frequency, at the same time emits radiation at the same frequency. The rate at which emission takes place is a function of temperature and frequency and is related to the Planck law which for a black body (BB) gives the following isotropic specific intensity I_{BB} :

$$I_{BB}(\mathbf{r}) = \frac{hv^3}{c^2} \frac{1}{e^{hv/KT} - 1} \quad (19)$$

where h and K are the Planck and Boltzmann constants, respectively, c is wave velocity, v the frequency, and T the temperature. For $hv \ll KT$ (valid at microwaves), $I_{BB} \approx KT/\lambda^2$ with λ the wavelength. It is easy to show that for the thermal emission specific intensity, it holds:

$$dI(\mathbf{r}, \hat{\mathbf{s}}) = +ds\Xi(\mathbf{r}, \hat{\mathbf{s}}) = +ds\kappa_a(\mathbf{r}, \hat{\mathbf{s}})I_{BB}(\mathbf{r}, \hat{\mathbf{s}}) \quad (20)$$

which shows that the emission power Ξ per unit volume, solid angle, and frequency of a medium is equal to the product κ_a times I_{BB} .

Adding all the contributions in Equations 17, 18, and 20, we get the *integrodifferential equation of radiative transfer*:

$$\frac{dI(\mathbf{r}, \hat{\mathbf{s}})}{ds} = \cdot [I(\mathbf{r}, \hat{\mathbf{s}})] = -\kappa_e(\mathbf{r}, \hat{\mathbf{s}})I(\mathbf{r}, \hat{\mathbf{s}}) + \int_{4\pi} \kappa_d(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}')I(\mathbf{r}, \hat{\mathbf{s}}')d\Omega' + \kappa_a \quad (21)$$

where the second term expresses the divergence of the specific intensity vector. Equation 21 can be also rearranged by introducing the so-called optical distance τ , defined by (Chandrasekhar, 1960)

$$\tau(\mathbf{r}, \hat{\mathbf{s}}) = \int_{4\pi} \kappa_e(\mathbf{r}, \hat{\mathbf{s}})ds \quad (22)$$

By using the definition of the scattering phase function in Equation 14, the RTE in Equation 21 is also given by

$$\frac{dI(\mathbf{r}, \hat{\mathbf{s}})}{d\tau} = -I(\mathbf{r}, \hat{\mathbf{s}}) + \frac{w(\mathbf{r}, \hat{\mathbf{s}})}{4\pi} \int_{4\pi} p(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}')I(\mathbf{r}, \hat{\mathbf{s}}')d\Omega' + [1 - w(\mathbf{r}, \hat{\mathbf{s}})]I_{BB}(\mathbf{r}) \quad (23)$$

At microwaves, it is useful to introduce the brightness temperature $T_B(\mathbf{r}, \hat{\mathbf{s}})$, defined as

$$I(\mathbf{r}, \hat{\mathbf{s}}) = \frac{K}{\lambda^2} T_B(\mathbf{r}, \hat{\mathbf{s}}) \quad (24)$$

where the brightness temperature is also defined as the product of the body emissivity times physical temperature. Through Equation 24, the RTE becomes

$$\frac{dT_B(\mathbf{r}, \hat{\mathbf{s}})}{d\tau} = -T_B(\mathbf{r}, \hat{\mathbf{s}}) + \frac{w(\mathbf{r}, \hat{\mathbf{s}})}{4\pi} \int_{4\pi} p(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}')T_B(\mathbf{r}, \hat{\mathbf{s}}')d\Omega' + [1 - w(\mathbf{r}, \hat{\mathbf{s}})]T(\mathbf{r}) \quad (25)$$

It is very often convenient to separate the total specific intensity into two terms, the *reduced incident* specific intensity $I_c(\mathbf{r}, \hat{\mathbf{s}})$ and *diffuse* specific intensity $I_d(\mathbf{r}, \hat{\mathbf{s}})$:

$$I(\mathbf{r}, \hat{\mathbf{s}}) = I_r(\mathbf{r}, \hat{\mathbf{s}}) + I_d(\mathbf{r}, \hat{\mathbf{s}}) \quad (26)$$

Substituting the previous equation into Equation 24, we obtain

$$\begin{cases} \frac{dI_r(\mathbf{r}, \hat{\mathbf{s}})}{d\tau} = -I_r(\mathbf{r}, \hat{\mathbf{s}}) \\ \frac{dI_d(\mathbf{r}, \hat{\mathbf{s}})}{d\tau} = -I_d(\mathbf{r}, \hat{\mathbf{s}}) + \frac{w(\mathbf{r}, \hat{\mathbf{s}})}{4\pi} \int_{4\pi} p(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}')I_d(\mathbf{r}, \hat{\mathbf{s}}')d\Omega' + J(\mathbf{r}, \hat{\mathbf{s}}) \end{cases} \quad (27)$$

where the *source function* $J(\mathbf{r}, \hat{\mathbf{s}})$ is given by the thermal term J_t and reduced incident term J_r :

$$J(\mathbf{r}, \hat{\mathbf{s}}) = [1 - w(\mathbf{r}, \hat{\mathbf{s}})]I_{BB}(\mathbf{r}) + \frac{w(\mathbf{r}, \hat{\mathbf{s}})}{4\pi} \int_{4\pi} p(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}')I_c(\mathbf{r}, \hat{\mathbf{s}}')d\Omega' + J_t(\mathbf{r}, \hat{\mathbf{s}}) \quad (28)$$

Note that from Equation 27, the reduced incident specific intensity is equal to the incident specific intensity reduced by the medium extinction coefficient κ_e .

The boundary conditions for the specific intensity at the surface S of medium containing random distribution scatterers may be various: (1) if no diffuse intensity is entering the medium, it holds $I_d(\mathbf{r}, \hat{\mathbf{s}}) = 0$ on S ; (2) if the incident intensity I_i is a collimated beam directed along $\hat{\mathbf{s}}_0$, then it holds $I_i(\mathbf{r}, \hat{\mathbf{s}}_0) = I_0 \delta(\hat{\mathbf{s}} - \hat{\mathbf{s}}_0)$ on S so that $I_r(\mathbf{r}, \hat{\mathbf{s}}_0) = I_0 \delta(\hat{\mathbf{s}} - \hat{\mathbf{s}}_0) \exp(-\tau)$; and (3) if the incident intensity I_i is a diffuse one itself, then it holds $I_i(\mathbf{r}, \hat{\mathbf{s}}_0) = I_0$ on S so that $I_r(\mathbf{r}, \hat{\mathbf{s}}_0) = I_0 \exp(-\tau)$.

Integral formulation and received power

The RTE for $I_d(\mathbf{r}, \hat{\mathbf{s}})$ in Equation 27 is first-order differential equation with respect to s . It is sometimes convenient to deal with the integral formulation especially for complex geometries. Supposing that at the boundary surface S in $\mathbf{r} = \mathbf{r}_0$ it holds $I_d(\mathbf{r}_0, \hat{\mathbf{s}}) = 0$ (see Figure 4), the general integral solution for $I_r(\mathbf{r}, \hat{\mathbf{s}})$ and $I_d(\mathbf{r}, \hat{\mathbf{s}})$ is

$$\begin{cases} I_r(\mathbf{r}, \hat{\mathbf{s}}) = I_i(\mathbf{r}_0, \hat{\mathbf{s}})e^{-\tau} \\ I_d(\mathbf{r}, \hat{\mathbf{s}}) = \int_0^s e^{-(\tau-\tau')} \left[\frac{w(\mathbf{r}', \hat{\mathbf{s}})}{4\pi} \int_{4\pi} p(\mathbf{r}', \hat{\mathbf{s}}, \hat{\mathbf{s}}')I_d(\mathbf{r}', \hat{\mathbf{s}}')d\Omega' + J(\mathbf{r}', \hat{\mathbf{s}}) \right] ds' \end{cases} \quad (29)$$

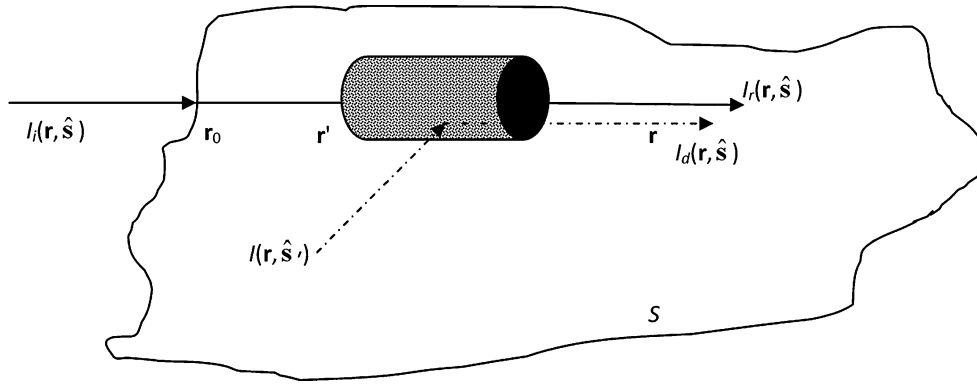
where $I_i(\mathbf{r}_0, \hat{\mathbf{s}})$ is the incident specific intensity and τ' is optical distance between $\mathbf{r}_0$ and $\mathbf{r}'$ with s the distance between $\mathbf{r}_0$ and $\mathbf{r}$.

In actual measurements of the specific intensity, it is necessary to take into account the characteristics of the receiver. The latter can be conveniently described by receiving equivalent area $A_e(\hat{\mathbf{s}}, \hat{\mathbf{s}}_0)$, measured in m^2 , whose maximum is supposed to be along $\hat{\mathbf{s}}_0$. If $I(\mathbf{r}, \hat{\mathbf{s}})$ is incident upon a receiver, placed at $\mathbf{r}$ and characterized by $A_e(\hat{\mathbf{s}}, \hat{\mathbf{s}}_0)$, the receiver power per unit frequency $P_R(\hat{\mathbf{s}}_0)$ (measured in W/Hz) is given by

$$P_R(\mathbf{r}, \hat{\mathbf{s}}_0) = \int_{\Delta\Omega} A_e(\mathbf{r}, \hat{\mathbf{s}}_0, \hat{\mathbf{s}})I(\mathbf{r}, \hat{\mathbf{s}})d\Omega \quad (30)$$

$$d\Omega = P_{Rr}(\mathbf{r}, \hat{\mathbf{s}}_0) + P_{Rd}(\mathbf{r}, \hat{\mathbf{s}}_0)$$

where $\Delta\Omega$ is the angular domain integration and the last term is due to the decomposition of I into I_r and I_d . The receiver equivalent area can be, for example, either a pencil beam (e.g., highly directive microwave antennas or optical detector) or a near-isotropic radiator (e.g., short dipole).



Radiative Transfer, Theory, Figure 4 Incident, coherent, and diffuse specific intensity in a volume V of random scatterers.

Radiative transfer of partially polarized radiation

The inclusion of polarization, while unnecessary for acoustic waves, may be important for electromagnetic waves. All electromagnetic waves in random media are, in general, partially polarized because, even though the incident wave is linearly polarized, the scattered wave is generally elliptically polarized and its polarization should randomly vary in time and space.

The scalar RTE can be generalized to partially polarized wave by replacing the scalar specific intensity with the *vector specific intensity* $\mathbf{I}(\mathbf{r}, \hat{\mathbf{s}})$, measured in $\text{W m}^{-2} \text{sr}^{-1} \text{Hz}^{-1}$, whose components are the Stokes parameters $I_h(\mathbf{r}, \hat{\mathbf{s}})$, $I_v(\mathbf{r}, \hat{\mathbf{s}})$, $U(\mathbf{r}, \hat{\mathbf{s}})$, and $V(\mathbf{r}, \hat{\mathbf{s}})$. The vector specific intensity can be defined starting from an elliptically polarized plane wave, representing an electrical field $\mathbf{E}(\mathbf{r}, \hat{\mathbf{s}})$:

$$\mathbf{E}(\mathbf{r}, \hat{\mathbf{s}}) = (E_h \hat{\mathbf{h}} + E_v \hat{\mathbf{v}}) e^{-j\mathbf{k} \cdot \mathbf{r}} \quad (31)$$

where $\hat{\mathbf{h}}$ and $\hat{\mathbf{v}}$ are the unit horizontal and vertical polarization vectors with E_h and E_v their complex amplitudes, j is the imaginary unit, and $\mathbf{k}$ is the propagation constant along $\hat{\mathbf{s}}$. The Stokes specific intensity vector $\mathbf{I}(\mathbf{r}, \hat{\mathbf{s}})$ is then given by

$$\mathbf{I}(\mathbf{r}, \hat{\mathbf{s}}) = \begin{bmatrix} I_v \\ I_h \\ U \\ V \end{bmatrix} = \begin{bmatrix} |E_v|^2 \\ |E_h|^2 \\ 2 \text{Re}(E_v E_h^*) \\ 2 \text{Im}(E_v E_h^*) \end{bmatrix} \quad (32)$$

Note the first two Stokes parameters are also called modified parameters, as the original ones introduced by Stokes were $I = I_v + I_h$ and $Q = I_v - I_h$. Moreover, sometimes the third term is divided by η , the medium intrinsic impedance.

An effective representation of the Stokes vector is through the so-called Poincaré sphere where in the Q , U , and V space, the I value is graphed with respect to the ellipticity and rotation angles (e.g., linear polarization shows a zero ellipticity angle and a 0° or 90° rotation angle if vertical or horizontal polarization is dealt with, respectively). For a *completely polarized* monochromatic wave,

it holds the following identity: $I^2 = Q^2 + U^2 + V^2$, whereas for partially polarized wave, it holds the inequality $I^2 > Q^2 + U^2 + V^2$. The *degree of polarization* δ_p is defined as the ratio between the square root of $(Q^2 + U^2 + V^2)$ and I^2 : If $\delta_p = 1$, the wave is *completely polarized*, whereas if $\delta_p = 0$, the wave is *completely unpolarized*. If $0 < \delta_p < 1$, the wave is said to be *partially polarized* and it corresponds to a point inside the Poincaré sphere (Ulaby and Elachi, 1990).

The vector analog of Equation 21 is the *vector RTE*:

$$\begin{aligned} \frac{d\mathbf{I}(\mathbf{r}, \hat{\mathbf{s}})}{ds} &= \kappa_e(\mathbf{r}, \hat{\mathbf{s}}) \mathbf{I}(\mathbf{r}, \hat{\mathbf{s}}) + \int_{4\pi} \kappa_d(\mathbf{r}, \hat{\mathbf{s}}') \mathbf{I}(\mathbf{r}, \hat{\mathbf{s}}') d\Omega' \\ &\quad + \kappa_a(\mathbf{r}, \hat{\mathbf{s}}) I_{BB}(\mathbf{r}) \end{aligned} \quad (33)$$

where $\mathbf{I}$ is the 4×1 Stokes column vector, $\kappa_e(\mathbf{r}, \hat{\mathbf{s}})$ is the 4×4 *extinction matrix*, $\kappa_d(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}')$ is the 4×4 *differential scattering matrix*, and $\kappa_a(\mathbf{r}, \hat{\mathbf{s}})$ is the 4×1 *absorption column vector*. Note that very often (and unfortunately), the matrix κ_d is also called phase matrix $\mathbf{P}$ even though it does not reflect the scalar definition of the phase function in Equation 14.

The *matrix* parameters in the vector RTE of Equation 31 can be related to the single-scattering property of each scatterer contained in the random medium. In a linear medium, we can relate the scattered field $\mathbf{E}_s(\mathbf{r}, \hat{\mathbf{s}})$ by a scatterer to the incident field $\mathbf{E}_i(\mathbf{r}, \hat{\mathbf{s}}')$ through the complex scattering 2×2 matrix $\mathbf{S}(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$:

$$\begin{aligned} \mathbf{E}_s(\mathbf{r}, \hat{\mathbf{s}}) &= \begin{bmatrix} E_{sv} \\ E_{sh} \end{bmatrix} = \begin{bmatrix} S_{vv} & S_{vh} \\ S_{hv} & S_{hh} \end{bmatrix} \begin{bmatrix} E_{ih} \\ E_{iv} \end{bmatrix} \\ &= \mathbf{S}(\hat{\mathbf{s}}, \hat{\mathbf{s}}') \mathbf{E}_i(\mathbf{r}, \hat{\mathbf{s}}') \end{aligned} \quad (34)$$

It is worth noting that the scattering matrix $\mathbf{S}$ can be expressed in any arbitrary orthonormal unit system, different from the one defined by the horizontal and the vertical polarizations as in Equation 34 (where usually the reference plane contains a preferential vertical axis and either the incident or scattered direction). An example is the

so-called scattering-plane orthonormal system which contains both the incidence direction and scattered direction where the polarization states are distinguished as either parallel or orthogonal to the scattering plane itself (Tsang et al., 2000). The scattering-plane reference system is convenient if the scatterer has symmetries.

By using the previous relation and Equation 32, we can easily show that, for a single scatterer, the scattered Stokes vector $\mathbf{I}_s(\mathbf{r}, \hat{\mathbf{s}})$ is related to the incident Stokes vector $\mathbf{I}_i(\mathbf{r}, \hat{\mathbf{s}}')$ by means of the so-called Stokes or Muller 4×4 matrix $\mathbf{L}(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$:

$$\mathbf{I}_s(\mathbf{r}, \hat{\mathbf{s}}) = \mathbf{L}(\hat{\mathbf{s}}, \hat{\mathbf{s}}') \mathbf{I}_i(\mathbf{r}, \hat{\mathbf{s}}') \quad (35a)$$

or explicitly:

$$\begin{bmatrix} I_{sv} \\ I_{sh} \\ U_s \\ V_s \end{bmatrix} = \begin{bmatrix} |S_{vv}|^2 & |S_{vh}|^2 & Re(S_{vh}^* S_{vv}) \\ |S_{hv}|^2 & |S_{hh}|^2 & Re(S_{hh}^* S_{hv}) \\ 2Re(S_{vv} S_{hv}^*) & 2Re(S_{vh} S_{hh}^*) & 2Re(S_{vv} S_{hh}^* + S_{vh} S_{hv}^*) - Im \\ 2Im(S_{vv} S_{hv}^*) & 2Im(S_{vh} S_{hh}^*) & Im(S_{vv} S_{hh}^* + S_{vh} S_{hv}^*) Re \end{bmatrix} \quad (35b)$$

The *differential scattering matrix* is then given, in analogy to the scalar case, as the weighted average of the Muller matrix $\mathbf{L}$ over the probability density function p_q :

$$\begin{aligned} \kappa_d(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}') &= N_q \int_q \mathbf{L}(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}', q_1, \dots, q_Q) p_q(q_1, \dots, q_Q) \\ &\quad d_{q_1} \dots d_{q_Q} \equiv \mathbf{P}(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}') \\ \kappa_d(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}') &= N_q \int_q \mathbf{L}(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}', q_1, \dots, q_Q) p_q(q_1, \dots, q_Q) \\ &\quad d_{q_1} \dots d_{q_Q} \equiv \mathbf{P}(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}') \end{aligned} \quad (36)$$

$$\begin{aligned} \kappa_d(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}') &= N_q \int_q \mathbf{L}(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}', q_1, \dots, q_Q) p_q(q_1, \dots, q_Q) \\ &\quad d_{q_1} \dots d_{q_Q} \equiv \mathbf{P}(\mathbf{r}, \hat{\mathbf{s}}, \hat{\mathbf{s}}') \end{aligned}$$

The previous matrix is also called the phase matrix $\mathbf{P}$, as indicated in the third term of Equation 36. The latter is essentially based on the assumption of *incoherent addition* of the diffuse Stokes parameters (Tsang et al., 2000).

In the same manner, the *extinction matrix* κ_e can be related to the complex scattering matrix $\mathbf{S}$. In particular, the optical theorem provides a useful relationship for computing the extinction cross section in terms of the field scattered in the forward direction, relative to the incident field. If the latter is t polarized, where $m = v$ or h , the p -polarized extinction cross section σ_{em} is

$$\sigma_{em}(\mathbf{r}, \hat{\mathbf{s}}) = \frac{4\pi}{k} Im[S_{mm}(\mathbf{r}, \hat{\mathbf{s}} = \hat{\mathbf{s}}', \hat{\mathbf{s}}')] \quad (37)$$

where Im is the imaginary-part operator and the scattering term S_{mm} is the mm scattering amplitude of the particle,

evaluated in the forward-scattering direction $\hat{\mathbf{s}} = \hat{\mathbf{s}}'$. The m -polarized extinction coefficient κ_{em} is given by the weighted average over the probability density function p_q :

$$\kappa_{et}(\mathbf{r}, \hat{\mathbf{s}}) = N_q \int_{\Delta q} \sigma_{et}(\mathbf{r}, \hat{\mathbf{s}}, q_1, \dots, q_Q) p_q(q_1, \dots, q_Q) d_{q_1} \dots d_{q_Q} \quad (38)$$

In the general case, for nonspherical particles, 4×4 extinction matrix κ_e is a non-diagonal matrix having the following general structure:

$$\kappa_e(\mathbf{r}, \hat{\mathbf{s}}) = \begin{bmatrix} -2Re(M_{vv}) & 0 & -Re(M_{vh}) & -Im(M) \\ 0 & -2Re(M_{hh}) & -Re(M_{hv}) & Im(M) \\ -2Re(M_{hv}) & -2Re(M_{vh}) & -Re(M_{vv} + M_{hh}) & Im(M_{vv} -) \\ -2Im(M_{hv}) & -2Im(M_{vh}) & -Im(M_{vv} + M_{hh}) & -Re(M_{vv}) \end{bmatrix} \quad (39)$$

Where

$$M_{mn} = \frac{j2\pi N_q}{k} \int_q S_{mn}(\mathbf{r}, \hat{\mathbf{s}}, q_1, \dots, q_Q) p_q(q_1, \dots, q_Q) d_{q_1} \dots d_{q_Q} \quad (40)$$

with $m, n = h, v$. The latter expressions can be derived from the theory of attenuation of the coherent wave through a random medium, adopting the so-called Foldy's approximation (Tsang et al., 1985). For the special case of a random medium with spherical particles, the 4×4 extinction matrix κ_e simplifies to a diagonal matrix with all identical elements $\kappa_{esph} = \kappa_{ehh} = \kappa_{evv}$:

$$\kappa_e(\mathbf{r}, \hat{\mathbf{s}}) = \begin{bmatrix} \kappa_{esph} & 0 & 0 & 0 \\ 0 & \kappa_{esph} & 0 & 0 \\ 0 & 0 & \kappa_{esph} & 0 \\ 0 & 0 & 0 & \kappa_{esph} \end{bmatrix} \quad (41)$$

Note that the absorption coefficient κ_a , usually due to gases, is an unpolarized quantity within the expression of the total extinction matrix κ_e . The emission vector of the random distribution of particles within the vector RTE is expressed through the absorption vector κ_a in Equation 31. Its expression might be fairly complicated in the general case and can be derived from the fluctuation-dissipation theorem for a single scatterer and the incoherent addition of Stokes parameters (Tsang et al., 1985). It can be shown that for random media of nonspherical particles, it holds

$$\kappa_a(\mathbf{r}, \hat{\mathbf{s}}) = \begin{bmatrix} \kappa_{a1}(-\hat{\mathbf{s}}) \\ \kappa_{a2}(-\hat{\mathbf{s}}) \\ \kappa_{a3}(-\hat{\mathbf{s}}) \\ \kappa_{a4}(-\hat{\mathbf{s}}) \end{bmatrix} \quad (42)$$

Where

$$\begin{aligned}
 \kappa_{a1}(\hat{s}) &= \kappa_{e11}(\hat{s}) - \int_{4\pi} [\kappa_{d11}(\hat{s}, \hat{s}') + \kappa_{d21}(\hat{s}, \hat{s}')] d\Omega' \\
 \kappa_{a2}(\hat{s}) &= \kappa_{e22}(\hat{s}) - \int_{4\pi} [\kappa_{d12}(\hat{s}, \hat{s}') + \kappa_{d22}(\hat{s}, \hat{s}')] d\Omega' \\
 \kappa_{a3}(\hat{s}) &= -2\kappa_{e13}(\hat{s}) - 2\kappa_{e23}(\hat{s}) + 2 \int_{4\pi} [\kappa_{d13}(\hat{s}, \hat{s}') + \kappa_{d23}(\hat{s}, \hat{s}')] d\Omega' \\
 \kappa_{a4}(\hat{s}) &= 2\kappa_{e14}(\hat{s}) + 2\kappa_{e24}(\hat{s}) - 2 \int_{4\pi} [\kappa_{d14}(\hat{s}, \hat{s}') + \kappa_{d24}(\hat{s}, \hat{s}')] d\Omega'
 \end{aligned} \tag{43}$$

where the elements κ_{emn} and κ_{dmn} are the mn elements of κ_e and κ_d , respectively, with $m, n = 1, 2, 3, 4$. Equations 42 and 43 state that the emission of the particle ensemble in the forward direction is related to the absorption by the particle ensemble in the backward direction.

Of course, the boundary conditions for the vector RTE need to be specified for the Stokes vectors as well. For example, in case of a planar dielectric interface, the Fresnel theory can be generalized so that the reflected Stokes vector $\mathbf{I}_r$ is given by

$$\mathbf{I}_r(\mathbf{r}, \hat{s}) = \mathbf{R}(\theta_i) \mathbf{I}_i(\mathbf{r}, \hat{s}) = \begin{bmatrix} R_v(\theta_i) & 0 & 0 & 0 \\ 0 & R_h(\theta_i) & 0 & 0 \\ 0 & 0 & \text{Re}(\Gamma_v \Gamma_h^*) & -\text{Im}(\Gamma_v \Gamma_h^*) \\ 0 & 0 & \text{Im}(\Gamma_v \Gamma_h^*) & \text{Re}(\Gamma_v \Gamma_h^*) \end{bmatrix} \tag{44}$$

where $\mathbf{R}$ is the *surface reflectance matrix* and $R_v(\theta_i) = |\Gamma_v(\theta_i)|^2$ and $R_h(\theta_i) = |\Gamma_h(\theta_i)|^2$ are the vertically and horizontally polarized, respectively, surface reflectance coefficients at the incident angle θ_i (satisfying the Snell law) with Γ_v and Γ_h , the Fresnel reflection field coefficients. For a Lambertian surface, the *surface reflectance matrix* is a matrix with all terms equal to zero except R_{11} and R_{22} both equal to 0.5.

Numerical and approximate solutions

The radiative transfer equation, in its scalar and vector form, is an integrodifferential equation which does not have analytical solutions, except in some special cases. Approximations and numerical techniques are usually adopted for solving the RTE (Chandrasekhar, 1960; Sobolev, 1975; Ishimaru, 1978; Tsang et al., 1985; Ulaby et al., 1986). Among these techniques, we enumerate (1) the *iterative solution* technique, (2) the *discrete-ordinate* technique and the *invariant imbedding* method, (3) the *invariant imbedding* technique, (4) the *two-flux* approximate solution, (5) the *Eddington* approximate solution, and (6) the *diffusion* approximate solution.

In the *iterative* solution, the radiative transfer equation is cast into an integral form and then solved iteratively to obtain the zero-, first-, and second-order solutions

(e.g., Tsang et al., 1985). This technique is applicable when the random medium is weakly scattering, that is, when the albedo w is small. The s solution is computed by ignoring scattering except for its contribution to extinction (the multiple-scattering term is disregarded). In principle, we can obtain an accurate solution by iterating many times. But, in practice, the iteration beyond the second order requires a substantial amount of computational efforts.

The *discrete-ordinate* eigen-analysis technique provides a fairly accurate numerical solution for the RTE and has been extensively used in literature (Chandrasekhar, 1960; Ishimaru, 1978). In this technique, the specific intensity and the phase matrix are first discretized into a finite number of directions by means of a quadrature, typically using a Fourier series expansion in the azimuth angle. It can be shown that for the incident plane wave case, all the Fourier components are independent of each other. Moreover, for the vertically and horizontally polarized waves, the first two terms of the Stokes vector are even functions and the last two terms are odd functions of the azimuth angle, respectively. Then, by means of the Gaussian quadrature typically, the resulting matrix differential equation for the discrete zenith angles is solved by eigen analysis. The discrete-ordinate technique is especially useful for plane-parallel media, but other methods, equally effective, should be also mentioned. Among these, the *invariant imbedding* (also called adding-doubling or matrix doubling in the scalar and vector RTE, respectively) is based on the general relationships between the reflection and transmission for a finite elementary slab and the formulation of integral equations relating them for a layered plane-parallel problem (Tsang et al., 1985).

For a one-dimensional (1D) problem, instead of using a series expansion technique or a discrete-ordinate approach, an approximate solution can be used in order to reduce the computational burden. The *two-flux theory* is, in this respect, very effective if the incident radiation is diffuse and the medium is dull (Kubelka and Munk, 1931). In this case, the RTE is reduced to two coupled differential equations expressed in terms of the forward F^+ and backward F^- fluxes. A drawback of the two-flux theory is that the equation coefficients are not constant and may depend on the angular characteristics of the intensity (they become constant only if the wave is completely diffused). In case of an incident collimated beam, the two-flux theory can be extended to the four-flux theory and successfully applied (Ishimaru, 1978).

For layered media, the *Eddington* solution is an effective approximate solution, based on the angular expansion of the specific intensity and the phase function in a series of Legendre polynomials with unknown coefficients and truncated at the first order (Eddington, 1916). Forward-scattering corrections of the basic solution can be used to improve the accuracy of the results in high albedo media (Joseph et al., 1976; Marzano, 2006).

In case of highly scattering media, the diffusion approximation can be used to reformulate the RTE. In this case,

the diffuse specific intensity is expressed as the average intensity plus a term which accounts for an increase in magnitude in the forward direction of the net flux (Ishimaru, 1978). Coupled with proper boundary conditions, the diffusion approximation takes the form of the fundamental steady-state diffusion equation. In case of *isotropic scattering*, the phase function can be assumed constant and the formulation further simplifies and can provide useful solutions when particles are much smaller than the wavelength (Menzel, 1966).

Limits and relation with the wave analytical theory

As already mentioned, the radiative transfer theory describes the power intensity balance of the transmission, absorption, and scattering processes and heuristically derived from considerations on energy conservation, as apparent in Equation 21. The RTE is based on the assumption that the correlation between the fields scattered by the different particles in the random medium is negligible and the addition of power, rather than the addition of fields, holds. However, the conventional RTE as in Equation 21 may be not valid for densely distributed media and fails to account for the backscattering enhancement effect which may be significant under certain circumstances (Kuga and Ishimaru, 1989).

Backscattering enhancement is a phenomenon in which a strong increase in the diffuse intensity can be observed in the backward direction when the random medium is illuminated by a plane wave, caused by the constructive interference of the two waves propagating in the opposite directions in the random medium. The intensity enhancement may be twice as large as the unenhanced intensity and the angular width may be much smaller than 1° . The backscattering enhancement cannot be explained by the radiative transfer theory as it does not include the phase correlation between waves. In order to explain these effects, we need to resort to the field theory. In fact, the same problem of the RTT can be cast in terms of field quantities and their statistical characterization. When this approach is followed, we can refer to the *analytical wave theory* (or multiple-scattering theory), whose general formulation is well summarized by the integral equation set of Twersky-Foldy (Twersky, 1964; Ishimaru, 1978).

The *analytical wave theory* generally starts from the Maxwell wave equation of a scalar field $\Psi(\mathbf{r})$, which can represent either a component of the electric or magnetic field or a pressure wave and describes the various processes of multiple scattering in a detailed way, usually resorting to the concise formalism of the diagram method. The general result of is that the total field $\Psi(\mathbf{r})$ is composed by (1) the incident wave, (2) multiply scattered waves involving chains of successive scattering through different scatterers, and (3) multiply scattered waves containing all the paths which go through a scatterer more than once. This general picture is consistent with that schematically shown in Figure 1. The *Twersky theory* adopts the so-called expanded representation which

retains only the previous groups (1) and (2) but disregards the effects due to group (3). When the number of scattering events becomes large, the errors associated to the Twersky approximation become very small.

Starting from the Twersky expanded representation, it is then possible to provide a statistical description of the random scalar field $\Psi(\mathbf{r})$, distinguishing between the coherent (or average) field $\Psi_c(\mathbf{r})$ and the diffuse (also incoherent or fluctuating) field $\Psi_d(\mathbf{r})$:

$$\psi(\mathbf{r}) = \psi_c(\mathbf{r}) + \psi_d(\mathbf{r}) \quad (45)$$

where $\Psi_c(\mathbf{r}) = \langle \Psi(\mathbf{r}) \rangle$ and $\langle \Psi_d(\mathbf{r}) \rangle = 0$ with the angle brackets indicating ensemble average. It can be shown that the total field intensity $\langle |\Psi(\mathbf{r})|^2 \rangle$, which can be decomposed into the sum of coherent intensity $|\langle \Psi(\mathbf{r}) \rangle|^2$ and diffusive intensity $\langle |\Psi_d(\mathbf{r})|^2 \rangle$ from Equation 45, satisfies the so-called Twersky-Foldy integral equations. The latter are, in their turn, consistent with the first-order smoothing approximation of the more rigorous Dyson and Bethe-Salpeter equations (e.g., Furutsu, 1975).

Considering that both the RTE in Equation 21 and the Twersky-Foldy integral equations describe the same multiple-scattering phenomenon, even though from two different points of view, a link between the two theories is expected (Ishimaru, 1975). Indeed, if $\Gamma_\Psi(\mathbf{r}, \Delta\mathbf{r})$ is the correlation function (or mutual coherence function) of the random field $\Psi(\mathbf{r})$ in a given position $\mathbf{r}$ and at a distance $\Delta\mathbf{r}$, it can be shown that it holds the following approximate equality (Furutsu, 1975):

$$\Gamma_\Psi(\mathbf{r}, 0) \cong \int_{4\pi} I(\mathbf{r}, \hat{s}) d\Omega \quad (46)$$

The previous relation indicates that the angular integral of the total specific intensity I (i.e., the radiation intensity) is proportional to the total field intensity $\langle |\Psi(\mathbf{r})|^2 \rangle$. Equation 46 can be generalized showing that the field correlation function Γ_Ψ is related to the Fourier transform of the specific intensity I . It shows that the description of multiple-scattering effects, based on the radiative transfer theory, does include information concerning the field quantities in the form of mutual coherence function (Ishimaru, 1978).

It is worth noting that if the particles are sparsely distributed, the extinction coefficient κ_e is linearly proportional to the particle number concentration N_q , as in Equation 13. This is a valid assumption if the fractional volume density is less than 1 %. Beyond 1 %, the extinction coefficient is no longer linearly proportional to the number concentration, but starts decreasing in a way dependent on the size parameter, dielectric constant, and particle volume fraction. This behavior in dense media has been demonstrated by an extensive experimental and theoretical analysis (Kuga and Ishimaru, 1989; Tsang et al., 1985). The previously mentioned Twersky model is relatively simple, but applicable only for small particles with respect to the wavelength. For volume fractions

higher than 10 %, the quasi-crystalline approximation (QCA) with either the Percus-Yevick pair correlation function or with the coherent potential approximation has shown to be effective solutions (Tsang et al., 1985).

Bibliography

- Chandrasekhar, S., 1960. *Radiative Transfer*. New York: Dover.
- Davison, B., 1958. *Neutron Transport Theory*. London: Oxford University Press.
- Eddington, A. S., 1916. On the radiative equilibrium of stars. *Monthly Notices of the Royal Astronomical Society*, **77**, 16–35.
- Elachi, C., 1987. *Introduction to the Physics and Technology of Remote Sensing*. New York: Wiley.
- Fung, A. K., 1994. *Microwave Scattering and Emission Models and Their Applications*. Norwood: Artech House.
- Furutsu, K., 1975. Multiple scattering of waves in a medium of randomly distributed particles and derivation of the transport equation. *Radio Science*, **10**, 29–44.
- Ishimaru, A., 1975. Correlation function of a wave in a random distribution of stationary and moving scatterers. *Radio Science*, **10**, 45–52.
- Ishimaru, A., 1978. *Wave Propagation and Scattering in Random Media*. New York: Academic, Vol. I and II.
- Ishimaru, A., and Kuga, Y., 1982. Attenuation constant of coherent field in a dense distribution of particles. *Journal of the Optical Society of America*, **72**, 1317–1320.
- Joseph, J. H., Wiscombe, W. J., and Weinman, J. A., 1976. The delta Eddington approximation for radiative flux transfer. *Journal of Atmospheric Sciences*, **33**, 2452–2459.
- Kubelka, P., and Munk, F., 1931. Ein beitrag zur optic der farbanstriche. *Zeitschrift für Technische Physik*, **12**, 593–603.
- Kuga, Y., and Ishimaru, A., 1989. Backscattering enhancement by randomly distributed very large particle. *Applied Optics*, **28**, 2165–2169.
- Marzano, F. S., 2006. Modeling antenna noise temperature due to rain clouds at microwave and millimeter-wave frequencies. *IEEE Transactions of Antennas and Propagation*, **54**, 1305–1317.
- Menzel, D. H. (ed.), 1966. *Selected Papers on the Transfer of Radiation*. New York: Dover.
- Schuster, A., 1905. Radiation through a foggy atmosphere. *Astrophysics Journal*, **21**, 1–22.
- Schwarzschild, K., 1906. Nachrichten königlichen gesellschaft wissenschaft. *Göttingen Mathematisch-Physikalische Klasse*, **195**, 41–64.
- Sobolev, V. V., 1975. *Light Scattering in Planetary Atmospheres*. New York: Pergamon.
- Sommerfeld, A., 1956. *Thermodynamics of Statistical Mechanics*. New York: Academic.
- Tsang, L., and Ishimaru, A., 1987. Radiative wave equations for vector electromagnetic propagation in dense nonsparse media. *Journal of Electromagnetic Waves and Applications*, **1**, 52–72.
- Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley.
- Tsang, L., Kong, J. A., and Ding, K.-H., 2000. *Scattering of Electromagnetic Waves: Theories and Applications*. New York: Wiley.
- Twersky, V. 1964. On propagation in random media of discrete scatterers. In *Stochastic processes in mathematical physics and engineering: Proceedings of a Symposium in Applied Mathematics of the American Mathematical Society*, New York, pp. 84–116.
- Ulaby, F. T., and Elachi, C. (eds.), 1990. *Radar Polarimetry for Geoscience Applications*. Norwood: Artech House.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1986. *Microwave Remote Sensing*. Norwood: Artech House, Vol. III.

- Wen, B., Tsang, L., Winebrenner, D. P., and Ishimaru, A., 1990. Dense medium radiative transfer theory: comparison with experiment and application to microwave remote sensing and polarimetry. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 46–59.

Cross-references

[Electromagnetic Theory and Wave Propagation](#)
[Fields and Radiation](#)
[Media, Electromagnetic Characteristics](#)
[Optical/Infrared, Radiative Transfer](#)
[Radiation, Multiple Scattering](#)
[Radiation, Polarization, and Coherence](#)
[Radiation, Volume Scattering](#)
[Radiative Transfer, Solution Techniques](#)
[Remote Sensing, Physics and Techniques](#)

RADIO-FREQUENCY INTERFERENCE (RFI) IN PASSIVE MICROWAVE SENSING

David Kunkee

The Aerospace Corporation, Los Angeles, CA, USA

Synonyms

Electromagnetic interference (EMI)

Definition

Radio-frequency interference (RFI) in passive microwave remote sensing occurs when anthropogenic (man-made) signals contaminate calibrated radiometric brightness temperature measurements of naturally occurring background thermal radiation resulting in anomalous measurements.

Introduction

Radio-frequency interference (RFI) to space-based passive microwave remote sensing systems became an important consideration after brightness temperature measurements from the Advanced Microwave Scanning Radiometer on NASA's Earth Observing System (AMSR-E) showed widespread contamination from anthropogenic emissions. The AMSR-E, part of NASA's Aqua mission, was launched on May 4, 2002, and was the first space-based radiometer since the scanning multichannel microwave radiometer (SMMR) (the SeaSat 7 mission ended in October 1987 and carried the last SMMR) to measure upwelling brightness temperatures near 6.8 GHz. Contamination (RFI) of brightness temperature measurements by SMMR radiometers at 6.6 GHz (ending in October 1987) was not significant over North America. However, beginning in 2002, substantial contamination of radiometric brightness temperatures near 6.9 GHz, from the AMSR-E, was found over North America as well as other populated regions of the world (Li et al., 2004). Contaminated brightness temperatures near 6.8 GHz, a spectral region important for remote sensing

of all-weather [sea surface temperature](#) and soil moisture retrievals, helped initiate the current research in RFI mitigation for passive microwave measurements and focus attention on protection of the RF spectrum for scientific uses.

Background

Passive microwave radiometers respond to the naturally occurring upwelling background radiation from the Earth's surface and atmosphere. The power contained in these emissions is a function of bandwidth and determined by the formula $P = kTB$, where k is Planck's constant, $1.038 \times 10^{-23} \text{ J}\cdot\text{s}^{-2}$, $T = T$ is the effective brightness temperature in Kelvin, and B is the effective RF bandwidth of the radiometer in Hz. When the radiometer responds to anthropogenic signals either within or outside of the radiometer's primary passband, the total RF power received by the radiometer increases; $P_T = kT_B B + P_{RFI}$ where P_{RFI} is mean power of the anthropogenic signal(s) during the radiometer's period of integration adjusted by the radiometer's antenna beam pattern and receiver passband response characteristics. As a result, the effective brightness temperature measured by the radiometer is larger than the scene brightness temperature $\delta T_B = \frac{P_{RFI}}{kB}$. This value represents the direct level of contamination.

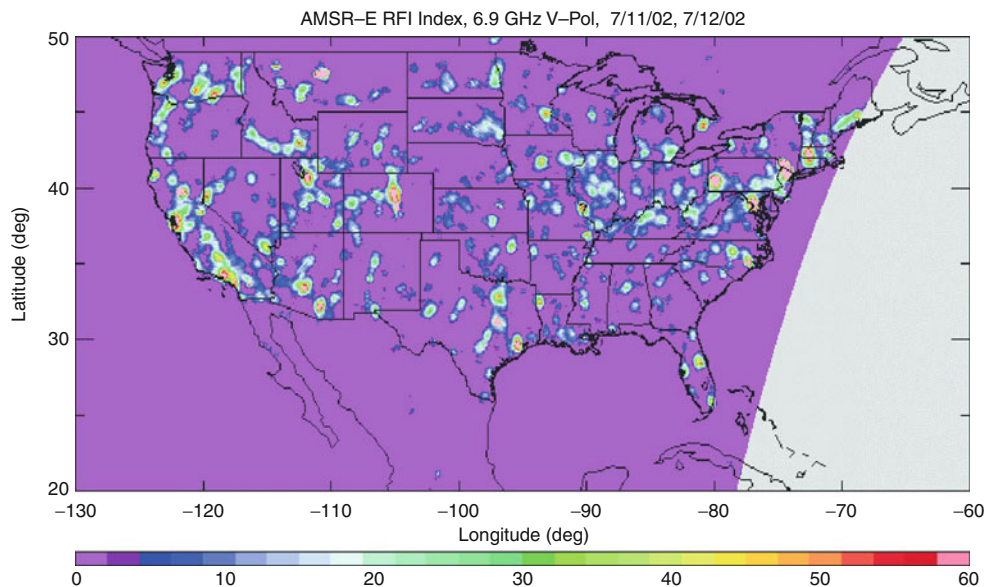
In general, the level of contamination, δT_B , is a function of radiometer bandwidth and strength of interfering signal. Larger bandwidth radiometers may tolerate higher levels of interference; however, they are sensitive to emissions over a wider range of frequencies. It is also important to

understand that out-of-band (OOB) signals also contribute to P_{RFI} . The OOB signals can be a significant consideration especially when strong anthropogenic emissions occur in the same spectral region that the radiometer is operating.

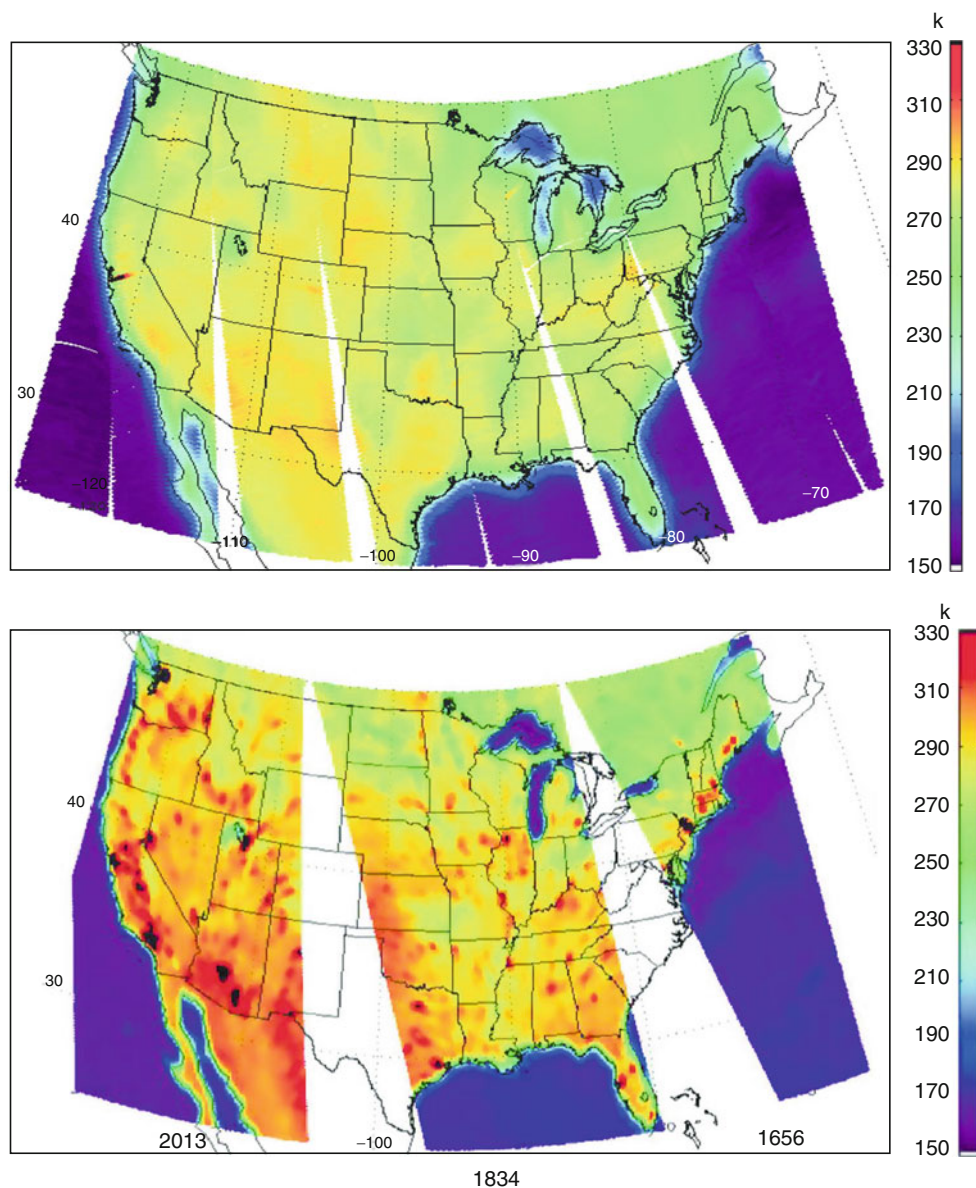
Another form of RFI that may impose errors on measured brightness temperatures occurs when strong OOB signals are present at the receiver (radiometer) input and impact the operating characteristics of the receiver. The OOB signals may otherwise be rejected by the receiver channelization filter; however, if the signals are of sufficient strength to impact operation of the receiver low noise amplifier (LNA), errors can occur in the measured data. These "indirect" effects from strong OOB signals may result in a net decrease or increase in measured brightness temperature depending on the nature of the response of the receiver front end and RF electronics.

Examples of RFI in several microwave bands

One of the most extensive examples of contaminated brightness temperatures can be seen in measurements of over North America from AMSR-E near 6.9 GHz (Figure 1). The data indicate that widespread utilization of this spectral region by other radio services (the spectrum segment is allocated to the fixed service (FS) and mobile service (MS) according to the current NTIA allocation chart (<http://www.ntia.doc.gov/osmhome/redbook/redbook.html>)) was causing substantial contamination to brightness temperatures measured by the AMSR-E vertically and horizontally polarized channels at 6.9 GHz.



Radio-Frequency Interference (RFI) in Passive Microwave Sensing, Figure 1 RFI is displayed as the perturbation from a zero-mean (natural emission) level determined by differencing 6- and 10 GHz brightness temperatures measured by AMSR-E. Perturbations of up to 50 K are common across the USA affecting more than 50 % of the total land area with RFI > 5 K. The pervasive nature of the interference resulted in reduced quality of soil moisture retrievals using AMSR-E because 6.9 GHz data could not be used (Li et al., 2003).

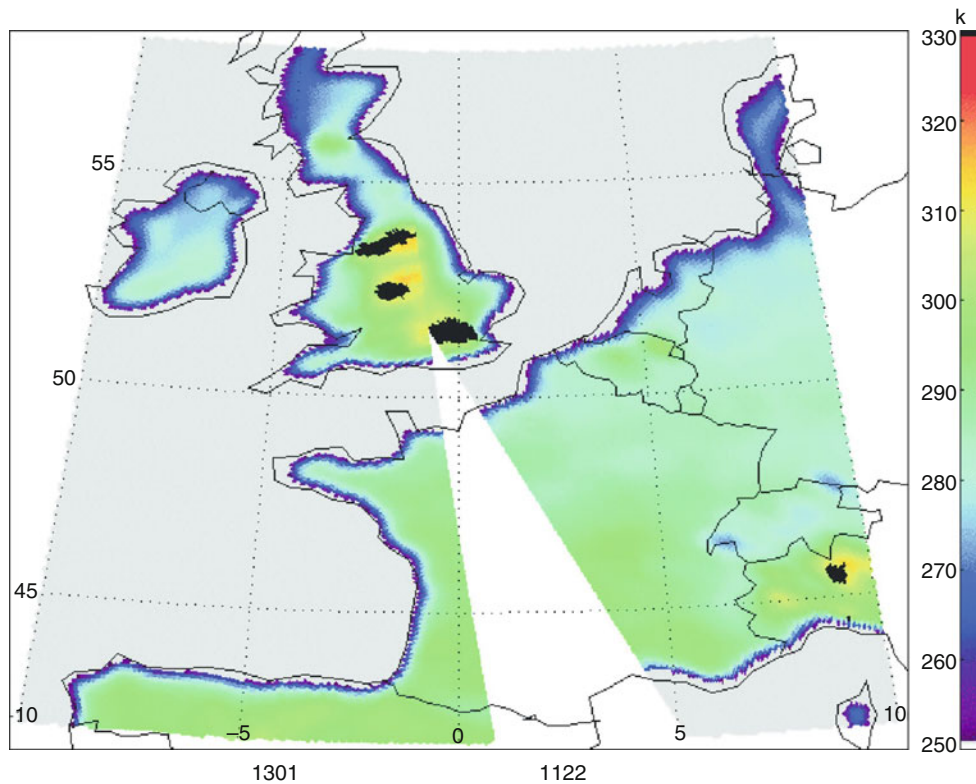


Radio-Frequency Interference (RFI) in Passive Microwave Sensing, Figure 2 (Top) Passive microwave brightness temperatures at 6.6 GHz from SMMR in October 1987. Essentially no RFI is seen over the North America in this period. (Bottom) Brightness temperature data at 6.9 GHz from AMSR-E in May 2003. The *black* spots represent high levels of anthropogenic emission primarily over regions of California and Arizona that saturate the AMSR-E radiometer. The *red* spots over most of the remaining areas of the USA represent contaminated brightness temperature measurements.

Contamination of background brightness temperatures is shown by departure from the mean difference of 6.9 and 10.7 GHz brightness temperatures in the figure. Positive values indicate RFI at 6.9 GHz. Contamination of brightness temperature measurements in the 6 GHz region was not found during the period of SMMR measurements from 1978 to 1987 (Njoku et al., 1980); brightness temperatures over North America were essentially contamination free as shown by data from SeaSat Nimbus 7 in contrast to AMSR-E data from 2002 (Figure 2). The red and black

spots in the lower part of the figure represent anthropogenic sources causing localized errors in measurements. In this example there is a high correlation of contaminated brightness temperatures with population centers.

RFI contamination to passive microwave brightness measurements has been observed in the 10 GHz region by the NRL WindSat radiometer, AMSR-E, and NASA's Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI). Two common types of occurrences have been noted. First, land-based anthropogenic



Radio-Frequency Interference (RFI) in Passive Microwave Sensing, Figure 3 Passive microwave brightness temperatures at 6.9 GHz from AMSR-E in June 2003 over France and the UK. The *black spots* represent saturation of the AMSR-E response indicating severe RFI. In this example occurrences are correlated with national boundaries.

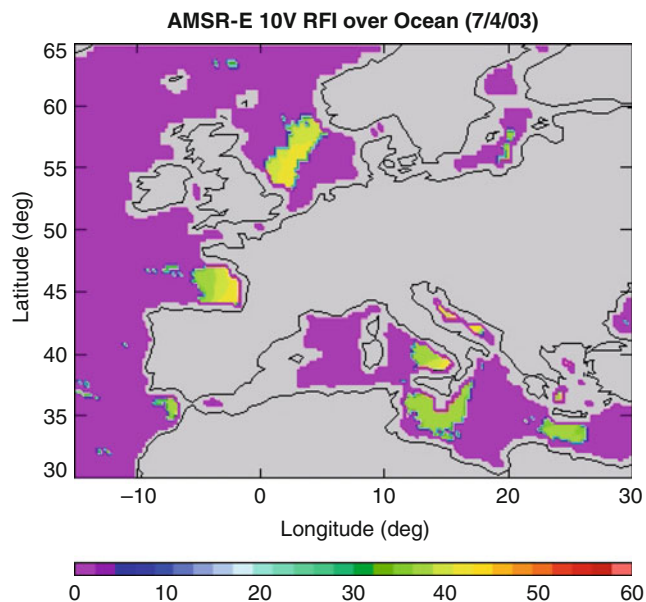
emissions have caused localized contamination of brightness temperature measurements (Figure 3). Second, reflections of geostationary broadcast transmissions to Earth from the ocean cause broad areas of contamination to measurements of brightness temperature (Figure 4). This contamination adversely impacts retrievals of *sea surface temperature (SST)* and sea surface wind (SSW) that rely directly on passive microwave brightness temperature data from the 10 GHz band.

Anomalies in sea surface wind retrievals near North America have also been attributed to RFI from geostationary broadcasts in the 18.6–18.8 GHz band segment. Flagging of potential anomalies in the WindSat 18.7 GHz data from NRL's retrieval algorithm is shown (Figure 5). Blue denotes fewer anomalies and red denotes a higher occurrence of anomalies. The first occurrence of these retrieval anomalies coincides with the activation of the DirecTV 10 geostationary broadcast satellite in early October 2007. The satellite operates within the band 18.3–18.8 GHz at 103° West longitude. In late July 2008, DirecTV 11 also went online at 99° West longitude with the same coverage area and down-link frequencies as DirecTV 10. The anomalies, found only during the descending phase of WindSat, are attributed to reflection of the satellite transmissions from the

ocean surface. The anomalies correlate precisely with the nationwide coverage maps of DirecTV 10 and DirecTV 11.

RFI mitigation techniques

After the AMSR-E brought RFI considerations to the forefront, development of RFI mitigation techniques has been under way by many researchers in passive microwave remote sensing. Techniques include frequency sub-banding (Gasiewski et al., 2002), detection of higher-order statistics (Ruf et al., 2006; Misra et al., 2008; De Roo et al., 2007; Piepmeier et al., 2008), and frequency-and time-domain methods (Johnson et al., 2006; Niamsuwan et al., 2005; Güner et al., 2007). Frequency sub-banding utilizes comparisons of measured brightness temperatures from each sub-band to determine if any single sub-band or multiple sub-bands are contaminated with RFI. Time-domain methods can be effective against pulsed emissions by utilizing short-duration blanking to eliminate the effect of the pulsed emissions. Higher-order statistics of the radiometric signal are generally sensitive to anthropogenic emissions (non-Gaussian) but are not sensitive to naturally occurring variations in scene brightness temperature. Therefore, measurements of higher-order statistics of the



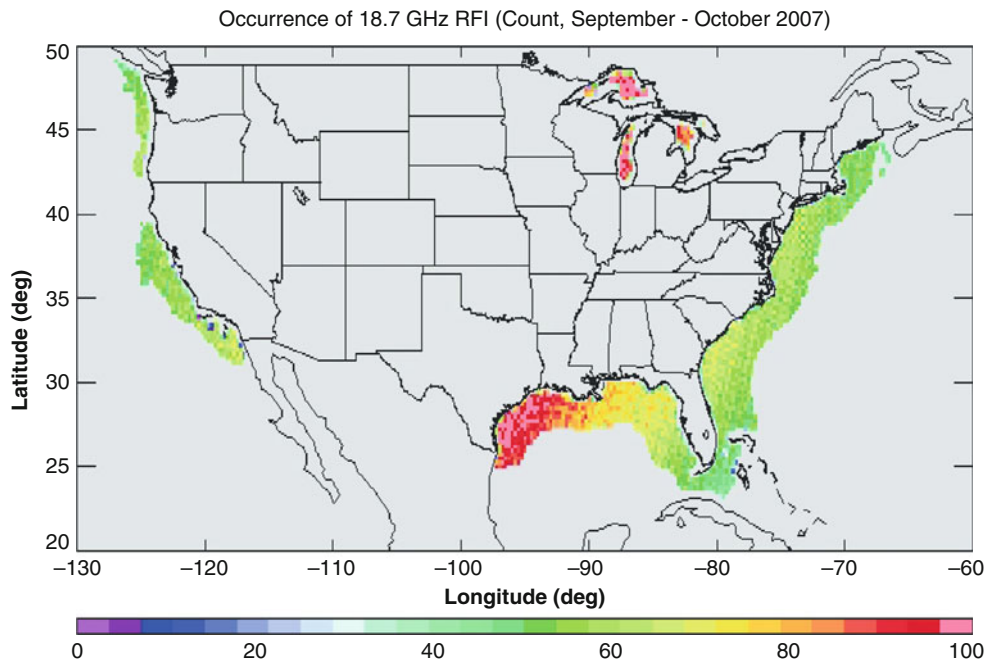
Radio-Frequency Interference (RFI) in Passive Microwave Sensing, Figure 4 Example of RFI incurred at 10 GHz from oceanic reflections of broadcasts in bands adjacent to those observed by AMSR (areas in green and yellow). In this example AMSR-E, operating in the EESS band 10.6–10.7 GHz, is experiencing > 40 K brightness temperature perturbations, a value that greatly exceeds minimum levels that degrade environmental models using SST data derived from AMSR-E (Courtesy Li Li US NRL).

signal such as the fourth moment (kurtosis) can be used to detect the presence of anthropogenic emissions and RFI and serve to enhance effective mitigation of its impact.

Although increasingly effective RFI mitigation techniques are under development, 100 % RFI mitigation will not be possible. Limited information from the scene brightness temperatures is lost in the presence of detectable anthropogenic emissions. The impacts and observations of RFI are also continuing to increase. Concurrent growth in communications and wireless markets are driving new applications of RF transceivers along with new developments and applications in passive microwave remote sensing that require improved levels of precision or additional frequency bands have resulted in increasing reports and concerns of RFI. For example, in the post SMMR radiometer era, 10 GHz space-based brightness temperature measurements from the TMI radiometer began in 1997, 6 GHz measurements were available from AMSR-E in 2002, and currently, there are two space-based radiometers operating at 1.4 GHz and another planned for 2014. These missions require new levels of radiometric precision and stability in a region of the spectrum that has many active users. As a result, there is increasing attention on protection of the RF spectrum for scientific uses.

Frequency allocations and regulation

In the USA, the National Telecommunications and Information Administration (NTIA) and Federal



Radio-Frequency Interference (RFI) in Passive Microwave Sensing, Figure 5 Occurrence of anomalous ocean wind retrievals near the North American shore. The anomalies began with the commencement of transmissions from a new geostationary broadcast satellite near 18.7 GHz on 1 October 2007 (Courtesy Li Li US NRL).

Communication Commission regulate the licenses for government and nongovernment uses, respectively, of the radio services. Licenses are issued in a manner consistent with the international allocation of frequencies as agreed by the International Telecommunication Union (ITU) and World Radio Conference (WRC) (ITU, 2001). WRCs are currently held every 2 years in order to address international frequency-coordination issues. One of the major tasks of the ITU is to provide recommendations for regulation of radio spectrum among the various radio services including the Earth Exploration Satellite Service (EESS) (passive microwave radiometry is classified as the Earth Exploration Satellite Service (EESS) – passive), radio astronomy service, satellite broadcast service, fixed service, and mobile service, for example. Recommendations are provided for EESS by committees at the US national and international levels called working party 7C (WP7C). The WP7Cs are composed of representatives that understand the regulatory and technical issues from agencies and organizations with an interest or mission to provide microwave remote sensing data for Earth-observation applications. One of the most important ITU recommendations relevant to EESS is ITU-R SA.1028 and SA.1029. These ITU recommendations suggest the maximum allowable anthropogenic emissions within sample bandwidths utilized for passive remote sensing and EESS to ensure interference-free operation. More information on the regulatory process related to scientific uses of the spectrum can be found in the Handbook of Frequency Allocations and Spectrum Protection for Scientific Uses (NRC, 2007).

Conclusion

Interference to space-based passive microwave remote sensing systems has continued to increase after the discovery of pervasive contamination of AMSR-E 6.9 GHz brightness temperature data. Transmissions originating with broadcast satellite systems operating in geosynchronous orbit near 10.7 and 18.7 GHz as well as land-based emissions in the 6-, 10-, and 18 GHz bands have been shown to impact passive microwave measurements. Over land, weak signals originating from point-to-point communication systems (fixed service) are sensed by space-based passive microwave radiometers observing near 10.65 and 18.6 GHz. Over ocean, data from NRL's WindSat show a negative impact to retrievals of sea surface temperature and ocean winds due to RFI at 10 GHz and recently due to RFI at 18.7 GHz near North America. In the past several years, the remote sensing and radio science communities have responded with research and improvements to RFI mitigation approaches; however, the effectiveness of the approaches is limited, and contaminated brightness temperature measurements result in quantifiable data loss even with effective RFI mitigation in place. From the regulatory perspective, allocations of radio spectrum to

specific radio services are determined by international agreement and documented by the ITU. Much of the spectrum is allocated to multiple radio services and sharing is required. In many cases, the requirements for noninterference to EESS systems by other radio services is challenging and not well understood. The ITU is also a source for recommendations concerning radio regulations and sharing criteria. The increasing pervasiveness of RFI to passive microwave measurements is a growing concern among remote sensing scientists. The Earth's naturally occurring microwave emissions represent a unique resource for monitoring and understanding our planet with direct application to weather forecasting and climate.

Bibliography

- De Roo, R. D., Misra, S., and Ruf, C. S., 2007. Sensitivity of the kurtosis statistics as a detector of pulsed sinusoidal RFI. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(7), 1938–1946.
- Gasiewski, A. J., Klein, M., Yevgrafov, A., and Leuskiy, V., 2002. Interference mitigation in passive microwave radiometry. In *IEEE International Geoscience and Remote Sensing Symposium*, 2002. IGARSS'02, Toronto, Vol. 3, pp. 1682–1684.
- Güner, B., Johnson, J. T., and Niamsuwan, N., 2007. Time and frequency blanking for radio frequency interference mitigation in microwave radiometry. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(11), 3672–3679.
- ITU, 2001. *Radio Regulations, International Telecommunications Union*. Geneva. <http://www.itu.int/net/home/index.aspx>.
- Johnson, J. T., Gasiewski, A. J., Güner, B., Hampson, G. A., Ellingson, S. W., Krishnamachari, R., Niamsuwan, N., McIntyre, E., Klien, M., and Leuski, V. Y., 2006. Airborne radio-frequency interference studies at C-band using a digital receiver. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(7), 1974–1985.
- Li, L., Njoku, E. G., Im, E., Chang, P. S., and St. Germain, K., 2004. A preliminary survey of radio-frequency interference over the U.S. in aqua AMSR-E data. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(1), 380–390.
- Misra, S., Ruf, C., and Kroodsma, R., 2008. Detectability of radio frequency interference due to spread spectrum communication signals using the kurtosis algorithm. In *IEEE International Geoscience and Remote Sensing Symposium*, 2008. IGARSS'08, Boston, Vol. 2, pp. 335–338.
- Niamsuwan, N., Johnson, J. T., and Ellingson, S. W., 2005. Examination of a single pulse blanking technique for radio frequency interference detection and mitigation. *Radio Science*, **40**(5), RS5–S03.
- Njoku, E., Stacey, J. M., and Barath, F. T., 1980. The seasat scanning multichannel microwave radiometer (SMMR): instrument description and performance. *IEEE Journal of Oceanic Engineering*, **OE-5**(2), 100–113.
- NRC, 2007. *Handbook of Frequency Allocations and Spectrum Protection for Scientific Uses*. Washington, DC: National Academies Press.
- Piepmeyer, P. M., and Knuble, J., 2008. A double detector for RFI mitigation in microwave radiometers. *IEEE Transactions on Geoscience and Remote Sensing*, **46**(2), 458–465.
- Ruf, C. S., Gross, S. M., and Misra, S., 2006. RFI detection and mitigation for microwave radiometry with an agile digital detector. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(3), 694–706.

RAINFALL

Ralph Ferraro
NOAA/NESDIS, ESSIC/CICS, College Park, MD, USA

Synonyms

Rain; Rainwater; Showers

Definition

Rainfall. A form of precipitation that reaches the Earth's surface in a liquid phase.

Introduction

The remote sensing of rainfall is a vital component to the integrated observing of precipitation on the Earth. While weather radars and rain gauges are the primary source of rainfall estimation in many areas, they are typically restricted to populated areas on the Earth and can only extend out over water bodies 150 km or so. Thus, satellites serve to fill in these huge data voids, especially over unpopulated regions and oceans. In this section, only satellite-based methods are considered (see *Water Resources; Global Earth Observation System of Systems (GEOSS)*). Additionally, this section refers to precipitation that falls at the surface in the form of liquid (see *Snowfall*).

A number of different methods are used to retrieve rainfall from satellites. The strength and weaknesses of the various methods that are described are summarized in *Table 1*, whereas details on the general classes of methods follow in the next three sections. Satellites (see *Observational Systems, Satellite*) that are used to estimate rainfall are generally categorized into LEO and GEO. Then, the various retrieval algorithms are typically classified on their observing spectrum (VIS, IR, PMW, AMW) or "multispectral" (i.e., use of one or more of these individual spectrums), whether they are objective or interactive (i.e., manual intervention) or whether they use multiple satellites or other information such as radar or gauges (e.g., "blended" algorithms).

Visible and infrared methods

Visible (VIS) and infrared (IR) techniques (see *Optical/Infrared, Radiative Transfer*) were the first to be conceived and are rather simple to apply, while at the same

time they show a relatively low degree of accuracy. On the other hand, GEO weather satellite VIS and IR imagers uniquely provide the rapid temporal update cycle (e.g., 30 min or less) needed to capture the growth and decay of precipitating clouds.

A complete overview of the early work and physical premises of VIS and thermal IR (10.5–12.5 μm) techniques is provided by Barrett and Martin (1981), while Kidder and Vonder Haar (1995) present some of the more recent results. The rainfall retrieval in these wavelengths is based on the fact that bright (optically thick) clouds are positively correlated with regions of convective rainfall (Woodley et al., 1971). On the other hand, clouds with cold tops in the IR imagery produce more rainfall than those with warmer tops (Scofield, 1987). However, the correspondence between cold tops and visible bright spots is far from perfect and is not always well correlated with surface rainfall (especially in stratiform rainfall regimes). *Figure 1* presents an example of typical VIS and IR signatures associated with different types of rainfall.

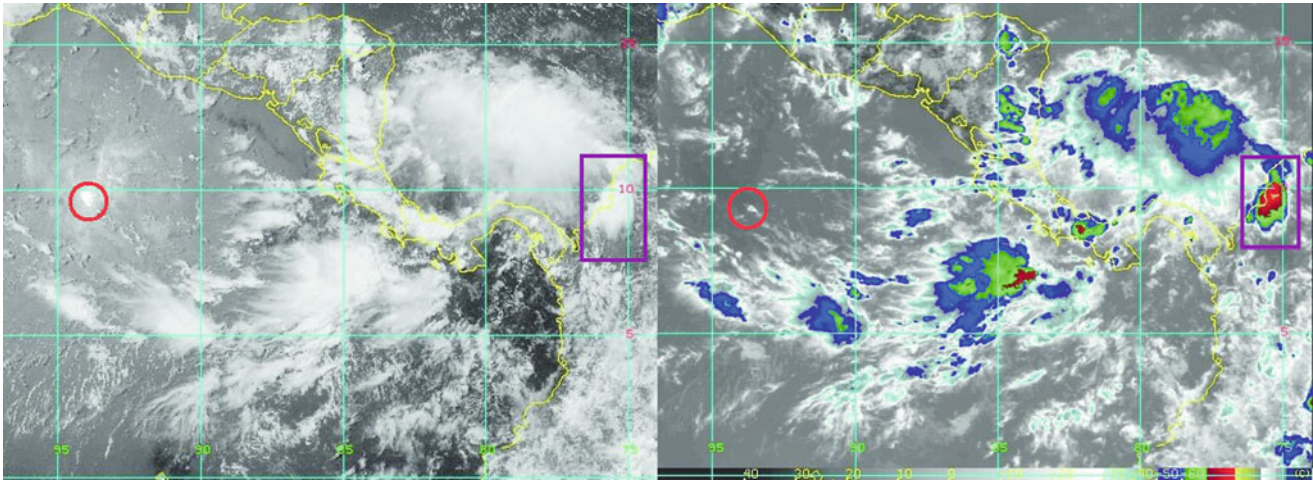
Various approaches have been developed to stress particular aspects of the sensing of cloud physics properties to settle differences between VIS and IR retrievals and measured rainfall. Following Barrett and Martin's classification, the rainfall estimation methods can be divided in the following categories: cloud indexing, bi-spectral schemes, life history, and cloud model based. In order to describe some methodologies that appear after Barrett and Martin's work, a new category that includes Numeric Weather Prediction (NWP)-adjusted schemes and multi-spectral algorithms has been added.

Cloud indexing methods

Cloud indexing techniques assign a rain rate level to each cloud type identified in the satellite imagery. The simplest and perhaps most widely used is the one developed by Arkin (1979) during the GARP (Global Atmosphere Research Program) Atlantic Tropical Experiment (GATE) on the basis of a high correlation between radar-estimated precipitation and fraction of the area colder than 235 K in the IR. The scheme, named the GOES Precipitation Index (GPI) (Arkin and Meisner, 1987), assigns these areas a constant rain rate of 3 mm h⁻¹, which is appropriate for tropical precipitation over 2.5° latitude $\times$ 2.5° longitude areas. The GPI is a standard for climatological

Rainfall, Table 1 Summary of satellite measurements and their attributes

Observation spectrum	Satellite type	Sensor examples	Strength	Weakness
Visible (VIS)	GEO	GOES imager	Cloud type	Cloud tops
	LEO	AVHRR	Cloud evolution	Indirect rain rate
Infrared (IR)	GEO	GOES imager	Cloud temperature	Cirrus contamination
	LEO	AVHRR	Cloud evolution	Indirect rain rate
Passive microwave (PMW)	LEO	SSM/I	Direct measure of rain, especially over ocean	Temporal sampling
		AMS-E		Indirect rain rate (land)
Active microwave (AMW)	LEO	TRMM PR	Direct measure of vertical structure of rain	Narrow swath width
		CloudSat CPR		Rain rate sensitivity



Rainfall, Figure 1 Comparison of VIS (left image) and IR (right image) from May 20, 2008 – 19:45Z over Central America. The red circle shows a feature with high reflectivity but high temperature (presumably low warm clouds with no rain associated), while the purple square shows a region with low temperature and very high reflectivity (large convective clouds). The last feature is related with higher rain rates.

rainfall analysis (Arkin and Janowiak, 1991) and is regularly applied and archived for climatological studies (see *Climate Monitoring and Prediction*).

A family of cloud indexing algorithms was developed at the University of Bristol. “Rain days” are identified from the occurrence of IR brightness temperatures (TB) below a threshold at a given location. The estimated rain days are combined with rain-per-rain day means that are spatially variable to produce rainfall estimations for extended periods (10 or more days). The most recent version of the Bristol technique uses variable IR rain/no-rain TB thresholds (Todd et al., 1995, 1999).

Bi-spectral methods

Bi-spectral methods are based on the very simple, although not always accurate, relationship between cold and bright clouds and high probability of precipitation, which is characteristic of cumulonimbus. Lower probabilities are associated to cold but dull clouds (thin cirrus) or bright but warm clouds (stratus). The RAINSAT technique (Lovejoy and Austin, 1979; Bellon et al., 1980) screens out cold but not highly reflective clouds or those that are highly reflective but have a relatively warm top. The number of false alarms of the pure IR techniques is reduced. The algorithm is based on a supervised classification trained by radar to recognize precipitation from both VIS brightness and IR brightness temperature.

Life-history methods

A family of techniques that specifically requires GEO satellite imagery are the life-history methods that rely upon a detailed analysis of the cloud’s life cycle, which is particularly relevant for convective clouds. An example is the Griffith-Woodley technique (Griffith et al., 1978). In this case, the increasing cloud area is treated different from

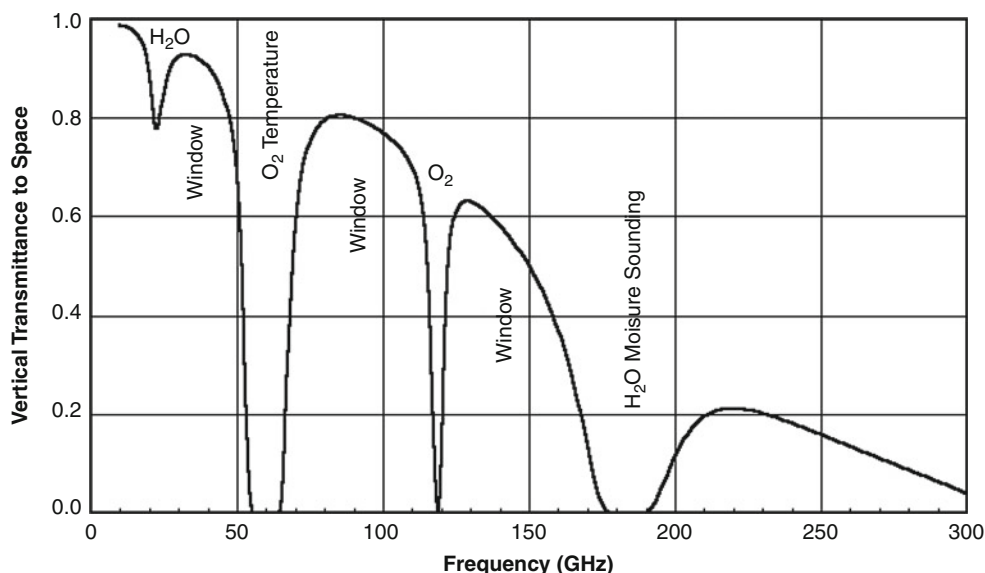
shrinking clouds. A major problem arises in the presence of cirrus anvils from neighboring clouds: they often screen the cloud life cycle underneath, leading to possible underestimates early in the day and overestimates toward the evening. Reasonable performances of this type of methods are obtained for deep convective storms, while contradictory results arise from their application to stratiform systems or weak convection (Amorati et al., 2000).

Cloud model-based techniques

Cloud model techniques aim at introducing the cloud physics into the retrieval process for a quantitative improvement deriving from the overall better physical description of the rain formation processes. A one-dimensional cloud model relates cloud-top temperature to rain rate and rain area in the Convective Stratiform Technique (CST) (Adler and Negri, 1988). Local minima in the IR brightness temperature are sought and screened to eliminate thin, nonprecipitating cirrus. To do so a slope parameter is calculated for each temperature minimum. Adler and Negri (1988) have established an empirical discrimination of thin cirrus in the temperature/slope plane using radar and visible imagery data. If the $T_{\min}$ and its slope fall to the left of the discrimination line, the $T_{\min}$ location is classified as thin cirrus (non-raining). A larger slope implies a more clearly defined minimum, corresponding to a thunderstorm.

NWP-adjusted and multispectral methods

Numerical weather prediction (NWP) outputs are being incorporated widely in satellite rainfall estimations. These correction factors range from moisture correction (Vicente et al., 1998) to orographic-driven precipitation. In the latter case, Digital Elevation Model (DEM) and low tropospheric winds are used to adjust the final



Rainfall, Figure 2 Vertical transmittance from the surface to space in the 15 °N annual atmosphere.

retrievals (Vicente et al., 2002). The Auto-Estimator technique (Vicente et al., 1998) proposed a moisture correction factor defined as the product of precipitable water (PW) in the layer from the surface to 500 mb times relative humidity (RH) (mean values between the surface and the 500 mb level) data. The PWRH factor decreases rainfall rates in very dry environments and increases them in very moist ones. The most recent version of this algorithm, called Hydro-Estimator, uses PW to adjust the power-law relation between cloud-top temperature and precipitation rates, while RH is applied to adjust the final rain rates (Scofield and Kuligowski, 2003).

The GOES Multispectral Rainfall Algorithm, GMSRA, (Ba and Gruber, 2001) combines multispectral measurements of the GEO satellites to estimate rainfall. One of the principal innovations of GMSRA relative to previous IR/VIS algorithms is that it combines several cloud properties used in a variety of techniques in a single and comprehensive rainfall algorithm. To be specific, the technique uses cloud-top temperatures as a basis of rain estimation (e.g., Arkin and Meisner, 1987; Vicente et al., 1998), and it utilizes the effective radii of cloud particles (e.g., Rosenfeld and Gutman, 1994) and spatial and temporal temperature gradients (e.g., Adler and Negri, 1988; Vicente et al., 1998) to screen out non-raining clouds.

Passive microwave methods

Unlike VIS and IR signals, microwave energy can penetrate clouds, in particular, cirrus clouds, and its signal has a strong interaction with precipitation-size drops and ice particles. This direct impact on microwave

measurements by hydrometeors allows for the quantitative detection of precipitation properties in the atmosphere as well as on the surface. It should be pointed out that passive MW (PMW) means naturally emitted radiation from the Earth's surface and atmosphere that interacts with clouds and precipitation and is measured by a radiometer onboard a satellite.

Most passive microwave radiometers launched to date operate in frequencies ranging from 6 to 190 GHz. At different frequencies, microwave radiometers observe different parts of the rain profile. Below 20 GHz, emission by precipitation-size drops dominates and ice particles above the rain layer are nearly transparent. Above 60 GHz, ice scattering dominates and the radiometers cannot sense the raindrops below the freezing layer. Both emission and scattering effects are important for frequencies between 20 and 60 GHz. In general, emission by liquid drops raises brightness temperature, while scattering by ice particles has the opposite effect. Besides the shift in dominating mechanism from emission to scattering with the increase of frequency, rain rate also plays a role by enhancing both the emission and the scattering signals. It is noted that the scattering by ice increases much more rapidly with frequency than scattering by liquid (Kidder and Vonder Haar, 1995). Other elements that have significant impacts on microwave radiation are shown in Figure 2. Window channels can measure down to the Earth's surface and are strongly influenced by surface properties (i.e., vegetation, soil moisture) (see *Land Surface Emissivity*). Other frequencies are sensitive to oxygen or water vapor/cloud droplets absorption. These microwave properties set the foundation for the development of rainfall estimation schemes.

There are two major categories in rainfall estimation using passive microwave radiometry: emission-based method and scattering-based method. They will be discussed in the following sections.

Emission-based methods

Emission-based rainfall algorithms are mostly applicable over ocean because water surfaces are relatively homogeneous and provide a cold background due to low emissivity. Below certain threshold rain intensity, emission dominates rather than scattering over water especially at lower frequencies. Brightness temperature (TB) increases rapidly with rain rate in this range and provides the basis for many emission-based precipitation retrieval methods. The primary disadvantage of emission-based techniques is that the microwave signal can saturate with moderate to heavy rain rates (which is frequency dependent), thus providing a strong nonlinear relationship between rain rate and TB.

The earliest efforts to estimate rain rate from satellite measurements used microwave data from the electrically scanning microwave radiometer (ESMR) on the Nimbus 5 and Nimbus 6 satellites and the scanning multichannel microwave radiometer (SMMR) on both Nimbus 7 and Seasat-A. Wilheit et al. (1977) employed a radiative transfer model to calculate the 19.35 GHz TB as a function of rain rate over ocean and used the 19.35 GHz measurements from ESMR-5 to retrieve rain rate. Weinman and Guetter (1977) utilized the weak polarization of 37 GHz radiances from ESMR-6 to discriminate convective rain over land from open water. Rain rate over water was expressed as a function of TB taking into consideration the polarization effects; a lower polarization was correlated to heavier rain rates. SMMR had five frequencies with both horizontal and vertical polarizations. Spencer et al. (1983) took advantage of the significant impact of raindrops on the thermal emission at lower frequencies (highest SMMR frequency is 37 GHz) and used multiple regression approach to relate SMMR TBs to rain rate.

The first of the special sensor microwave/imager (SSM/I) series was launched in 1987 and was followed by five more successful SSM/I instruments. The SSM/I has seven vertically and horizontally polarized channels and four frequencies at 19.35, 22.235, 37.0, and 85.5 GHz. This is a well-studied instrument for rainfall estimation. Ferraro and Marks (1995) developed an SSM/I algorithm linking oceanic rain rate with liquid water path (Weng and Grody, 1994) using ground-based radar measurements. The retrieval of liquid water path relies on the emission signatures at 19 and 37 GHz. This ocean scheme is part of the operational SSM/I rainfall algorithm as introduced in the next section.

Mugnai and Smith (1984) were the first to combine cloud model with radiative transfer model for estimating TBs from convective clouds. Studies along this line of thinking followed (Olson, 1989; Smith et al., 1994). The Goddard profiling algorithm (GPROF) is one of the

inversion algorithms that computes radiative transfer based on output from cloud models (Kummerow et al., 1996, 2001). GPROF is applied to the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) for the retrieval of surface rain rate and rainfall profile. A large a priori database of atmospheric profiles and TB at the TMI frequencies is constructed using output of two cloud-resolving models and a one-dimensional radiative transfer model. Then the algorithm utilizes a Bayesian inversion approach to derive instantaneous rain rate and rainfall profile by a weighted summation of all the profiles in the database. The weight is determined by how close each TB vector in the database resembles the observed vector. GPROF is also the primary rainfall algorithm being applied to the Advanced Microwave Sounding Radiometer-Earth Observing System (EOS) (AMSRE) (Wilheit et al., 2003).

Bauer et al. (2005) proposed a scheme for retrieving precipitation profiles over all surfaces by using microwave frequencies both in atmospheric windows between 18 and 150 GHz and in oxygen absorption complexes or sounding channels near 50–60 and 118 GHz. Hydrometeor profiles are created by applying cloud and convection schemes using information from European Centre for Medium-Range Weather Forecasts (ECMWF) short-range forecasts. A radiative transfer model and one-dimensional variational (1D-Var) retrieval framework are employed to achieve optimum hydrometeor profile retrievals.

Scattering-based methods

Due to the higher but more varying emissivity of the land surface, the only reliable means of detecting rainfall over land is by isolating depressed TBs as a result of scattering by millimeter-sized ice particles that exist in most rain clouds (see *Optical/Infrared, Scattering by Aerosols and Hydrometeors*). Since the signal being captured is a result of ice particles instead of raindrops, the scattering-based rainfall estimation is an indirect measure of rainfall, as it relates the magnitude of the scattering near the freezing layer to surface rainfall. The launch of the SSM/I in 1987 provided the first opportunity to retrieve rain rate through scattering at higher frequency (85 GHz). Meteorological satellites launched since then have included channels at or above 85 GHz that are sensitive to scattering, including more recent sensors that contain measurement channels at or above 150 GHz. Thus, these measurements have been exploited to develop scattering-based rainfall algorithms.

Spencer et al. (1989) first introduced the concept of polarization-corrected brightness temperature (PCT) which is a linear combination of the vertically and horizontally polarized brightness temperatures at 85.5 GHz. The coefficients in PCT can be adjusted so that PCT is only sensitive to the scattering in the upwelling microwave radiation but not to other atmospheric and land surface conditions including the contrast of land versus ocean.

An operational SSM/I rain rate product generated at Fleet Numerical Meteorology and Oceanography Center (FNMOC) (Ferraro, 1997) combines a scattering-based global algorithm with the abovementioned SSM/I emission-based ocean algorithm (Ferraro and Marks, 1995). The global algorithm calculates a scattering index (SI) (Grody, 1991) which is the difference between the vertically polarized 85.5 GHz observation and an estimate of the non-scattering contribution from this channel. The latter is a linear combination of the observations at lower channels (19.35 and 22.235 GHz). The SI algorithms are calibrated with ground-based radar measurements to produce instantaneous rain rate for both ocean and land (Ferraro and Marks, 1995). The emission-based algorithm complements the SI algorithm by detecting rain systems over ocean that has little or no ice scattering. A decision tree is included in the algorithm to screen out surface scattering sources such as snow cover and desert. This widely used algorithm is also adapted into GPROF land rain rate algorithm as described by McCollum and Ferraro (2003). The relationships between rain rate and 85 GHz brightness temperature are recalibrated using TRMM Precipitation Radar (PR) data, and a new procedure is developed to estimate convective rainfall.

Zhao and Weng (2002) took advantage of the highly scattering nature of 89 and in particular 150 GHz radiances from the Advanced Microwave Sounding Unit-B (AMSU-B) and retrieved ice water path (IWP) using scattering parameters measured at these two channels. The derived IWP is then converted into the surface rainfall rate (RR) through an IWP and rainfall rate relationship developed from cloud model results (Weng et al., 2003). This rain rate product is being operationally generated at National Environmental Satellite, Data, and Information Service (NESDIS) of National Oceanic and Atmospheric Administration (NOAA). This algorithm has also been applied to Microwave Humidity Sounder (MHS) with some modification since AMSU-B and MHS have very similar channels. Vila et al. (2008) added an emission-based component to this rainfall algorithm to account for oceanic rain systems that have little or no ice in them. Figure 3 shows this rain rate product and some of the corresponding TBs that have the emission and scattering signals of rain systems.

Active microwave methods

In contrast to passive microwave radiometers, active microwave sensors provide their own source of microwave radiation. Short pulses are transmitted at microwave frequencies; the time delay and strength of the returned echo gives the distance and intensity of the rain from space. Radar is also able to discriminate in range, that is, in altitude from space. Hence, radar measures fine-scale and vertical distribution of rainfall. In order to provide good horizontal resolution and minimize surface clutter effects for off-nadir incidence without the use of large antennas (greater than 2 m), only frequencies greater than

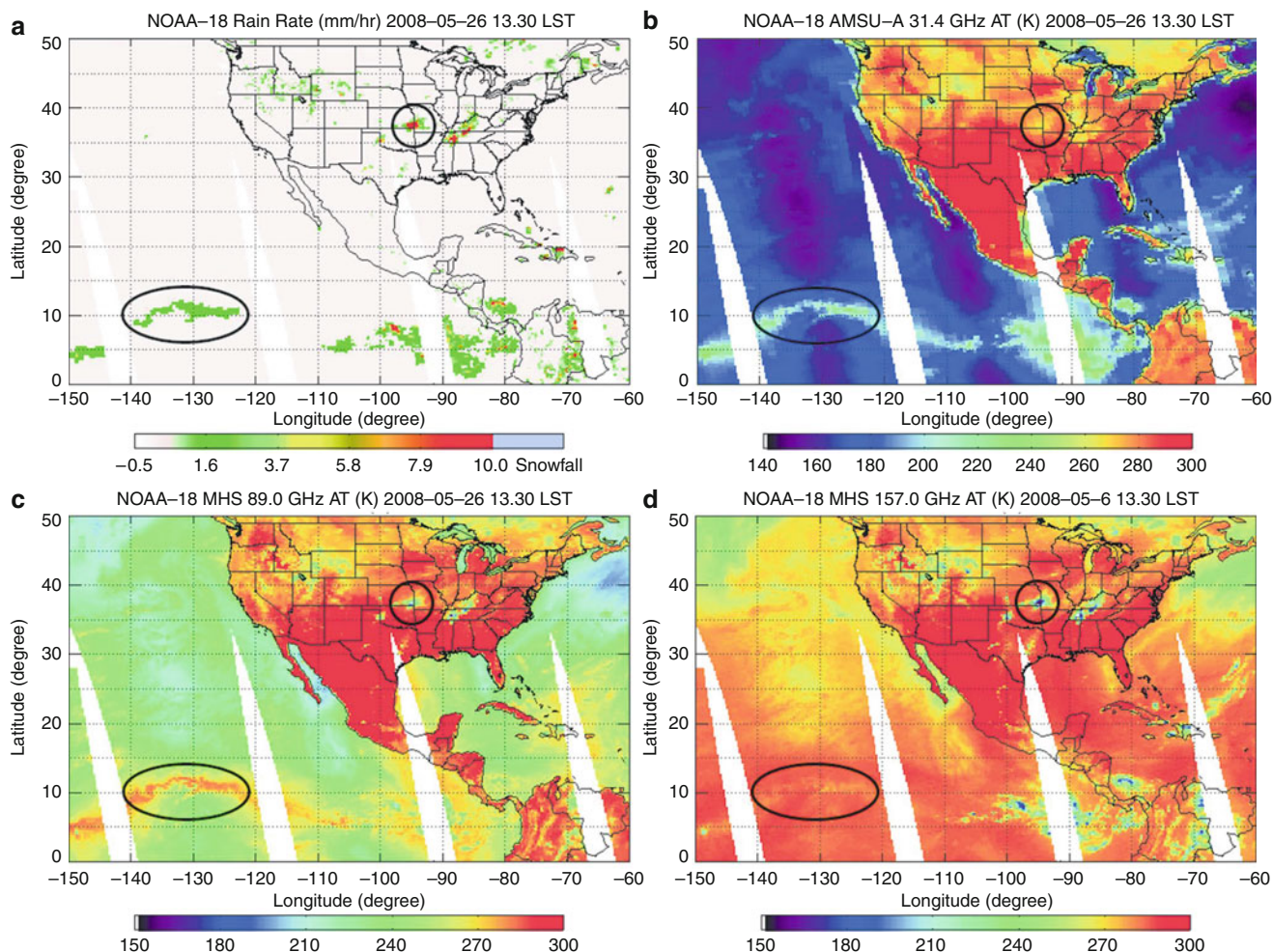
10 GHz are practical (Fujita and Satake, 1997). At such frequencies, attenuation is significant in moderate to heavy rain and affects the radar reflectivity measurement near the surface where rain information is most desired. Over the rain depth of kilometers, the two-way path-integrated attenuation (PIA) becomes significant even at moderate rain rates and leads to a bias in rain rate estimate, unless reflectivity is corrected for attenuation effects. Such a bias increases with frequency as a result of an increase in PIA.

Techniques have been developed to remove the effects of attenuation when they become significant. The study by Hitschfeld and Bordan (1954) was one of the first to propose a method for directly correcting measured reflectivities for attenuation. After correcting attenuation effects, estimates of non-attenuated reflectivity (Z) are related to rain rate (R) by using Z - R relations. The Hitschfeld-Bordan (HB) solution gives reasonable estimates if the attenuation is small and the radar is well calibrated. When these conditions are not met, the HB solution can become unstable. Numerous retrieval methods have been developed that incorporate this early work along with the more recent technique of constraining the retrievals using PIA derived from surface reference technique (SRT) (Iguchi and Meneghini, 1994; Marzoug and Amayenc, 1994; Amayenc et al., 1996). SRT relies on a reference measurement of the radar return from the surface in rain-free conditions. For example, if the surface returns for a particular incidence angle and surface types are fairly stable, then the difference in surface returns for rain-filled and rain-free radar beams can be ascribed to the effects of attenuation by precipitation. The SRT method provides a reasonable way to monitor PIA over the ocean. Over land, particularly at near-nadir incidence where the surface scattering cross section is usually highly variable in both space and time, the method is less reliable (Meneghini et al., 2000).

For cm-wavelength radars (such as the TRMM radar), attenuation corrections are relatively modest for most rains. For mm-wavelength radars (such as the radar on CloudSat), attenuation is strong and often becomes a dominant factor responsible for systematic reflectivity differences. Another example of rainfall retrieval technique from spaceborne radar is to use the attenuation directly caused by rainfall (Matrosov, 2007). This method takes advantage of high attenuation in rain and low variability of non-attenuated reflectivities and uses estimates of height derivatives of attenuated reflectivities. These estimates are then related to rain rate. Early result from a 94 GHz CloudSat radar indicates a possibility of rain rate profile retrieval (Matrosov, 2007; Mitrescu et al., 2008).

Current Sensors

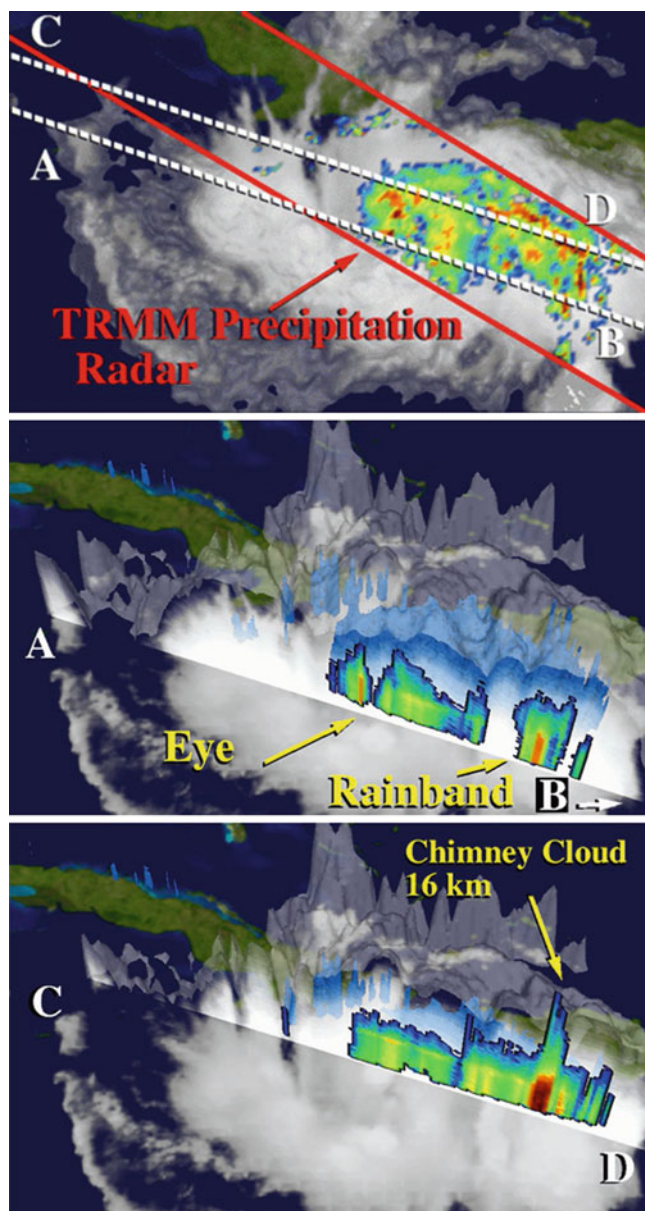
- (a) TRMM PR In orbit since 1997, the Precipitation Radar (PR) on the TRMM satellite (Tropical Rainfall Measurement Mission) is the first instrument designed to measure rain from space (Kummerow et al., 1998).



Rainfall, Figure 3 NOAA-18 rain rate product (NOAA/NESDIS) and antenna temperatures. (a) Rain rate and antenna temperature at (b) AMSU-A 31.4 GHz, (c) MHS 89 GHz, and (d) MHS 157 GHz. The black circle in the images indicates a rain system dominated with ice cloud and characterized with gradual increase in scattering signal from 31.4 to 157 GHz and hence decreasing in antenna temperatures. The scattering-based portion of the NESDIS rain rate algorithm captures this rain system very well. The black oval shows a warm rain system in the tropical Pacific Ocean that was detected by the emission-based portion of the NESDIS algorithm. The emission signal is visible in 31.4 and 89 GHz but weak in 157 GHz images. Comparing with the nonprecipitating ocean area, the antenna temperatures in the raining zone are warmer at 31.4 and 89 GHz but have little difference at 157 GHz.

TRMM's PR provides direct, fine-scale, and vertical distribution of precipitation. PR is a cross-track radar that surveys a swath of 220 km at a horizontal resolution of 4.3 km. It gives the vertical distribution of rain, with the height resolution of 250 m. TRMM is in a low orbit, 402.5 km above the Earth's surface. Its subtrack covers latitudes of $\pm 35^\circ$. The PR rain retrieval algorithm (Iguchi et al., 2000) uses a hybrid of the HB and SRT methods to correct for attenuation. When the SRT is reliable, the radar-rain rate relations are adjusted so that the HB path-attenuated estimate is equal to that of SRT. When the SRT is unreliable, the HB method is usually used along with radar-rainfall relation derived from ground-based disdrometer measurements, conditioned on rain type, to provide vertical profiles of rain rate. Figure 4

illustrates an example of PR's ability to probe vertical rain structure. On September 28, Hurricane Lili was southeast of Cuba, but winds in the tropical storm were blowing at a steady 45 kts. Figure 4a shows PR estimated rain in colors: clouds are shown in white; light rain is shown in blue; extreme rain rates in red. Figure 4b shows the cone-shaped eye, newly developed and surrounded by moderate intensity rain clouds. Also shown is a prominent rain band feeding in toward the storm's inner core. Figure 4c reveals towering clouds that helped power Lili's growth, called "chimney clouds." The chimneys contain intense updrafts that release copious amounts of heat energy inside the storm. Such three-dimensional information on precipitation can only be measured by a spaceborne radar.



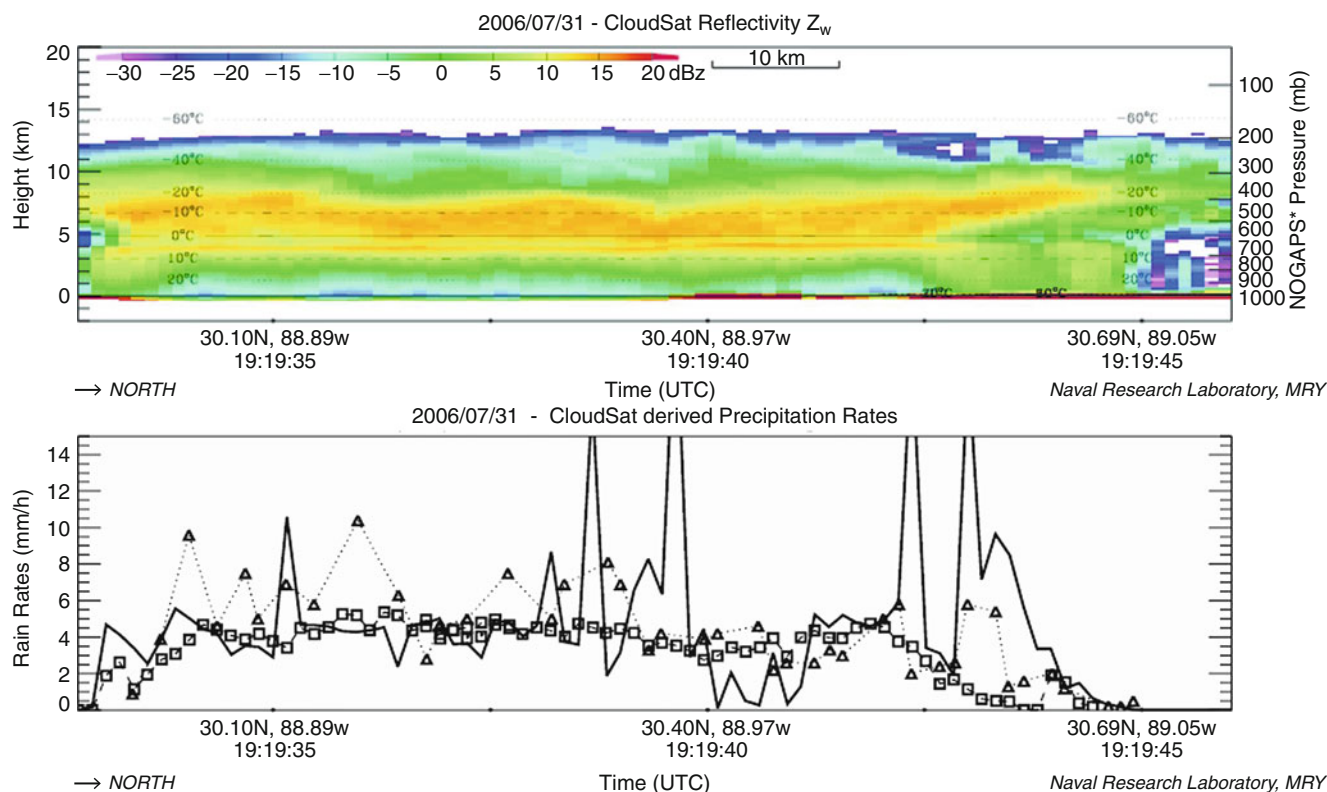
Rainfall, Figure 4 TRMM PR swath (red lines) and rain rates for Hurricane Lili on 28 September 2002 overlaid on a cloud image (top). PR vertical cross sections for line A–B (middle) and line C–D (bottom) as indicated in top image. The color shading in all images indicates the relative rain intensities (Image courtesy of NASA/GSFC).

- (b) CloudSat Launched in April 2006, CloudSat is the first millimeter-wavelength cloud radar in space (Stephens et al., 2002). The Cloud Profiling Radar (CPR) aboard CloudSat is a 94 GHz near-nadir-pointing radar with 500 m vertical resolution. The design concept of CPR is to provide vertical cross section of cloud liquid and ice water content and particle size. At the altitude of 750 km, the CPR provides vertical

profiles of radar reflectivity. Although the main objective of CPR is to provide global information on clouds, it also resolves many precipitation systems. It has been suggested that CloudSat data can be used to estimate light rainfall (L'ecuyer and Stephens, 2002). Early results have shown the potential of 94 GHz cloud radar for the detection of precipitation (Haynes et al., 2007) and rain rate profile retrievals (Matrosov, 2007). Figure 5 shows an example of CloudSat rain profiling retrieval (Mitrescu et al., 2008). Figure 5a shows the CloudSat reflectivity on 31 July 2006. The quasi-uniform cloud layer extends up to 13 km; the bright band is identifiable at 4 km. Figure 5b shows the CloudSat retrieved rain rates from two different retrieval algorithms and a ground radar measurement: solid line is retrievals by Mitrescu et al. (2008), squares are retrievals by Matrosov (2007), and triangles are rain rate estimates from the NEXRAD radar KLIX located in New Orleans. Although overall, the retrieved rain rates are comparable, there are noticeable differences.

Future sensors

- (a) GPM-DPR GPM (Global Precipitation Measurement) is a joint US-Japan mission designed to extend TRMM's observations of precipitation to higher latitudes, with more frequent sampling. The GPM core satellite will carry a dual-frequency, cross-track scanning, Ku/Ka-band (13.6/35.5 GHz) precipitation radar, i.e., the DPR (Iguchi et al., 2002). Anticipated advantages of the DPR over the single-frequency PR include enhanced sensitivity at light rain rates, information on the particle size distribution in rain and snow, and improvements in the identification of vertical layers consisting of mixed phase, frozen, and liquid particles. Particle size estimation follows from Mie scattering effects in that the difference in radar reflectivities (in dB) at the two frequencies provides an estimate of the mean particle size of the distribution (Doviak and Zrnica, 1984; Meneghini et al., 1989, 1992; Kuo et al., 2004). This dual-frequency measurement allows for recovery of at least 2 of 3 parameters needed to describe the bulk distribution of the generalized raindrop size distribution (DSD). Such measurements are important because increased knowledge of DSD factors improves the retrievals of rain rate.
- (b) EarthCARE EarthCARE is a joint Europe-Japan mission with a spaceborne 94 GHz Cloud Profiling Radar (CPR) and is scheduled to be launched in 2012. CPR on EarthCARE will have better sensitivity to cloud detection than CloudSat. One other important feature of CPR/EarthCARE will be its Doppler capability which is expected to provide vertical Doppler velocity profiles of cloud echoes, which are important for cloud physics and drizzle detection (Ohno et al., 2007).



Rainfall, Figure 5 CloudSat reflectivity in dBZ (*top*) and derived precipitation rates in mm h^{-1} (*bottom*) for 31 July 2006 for a short cross section through a storm system over the coastline of Mississippi. The upper image has the surface temperature and pressure contoured to give an indication of the height of the freezing level. The lower image shows the corresponding rain rate from two different CloudSat retrievals (*squares and triangles*) and surface radar-derived rain rates (*solid line*) (The image is courtesy of J. Turk and T. Mirtrescu, Naval Research Laboratory, Monterey, CA, USA).

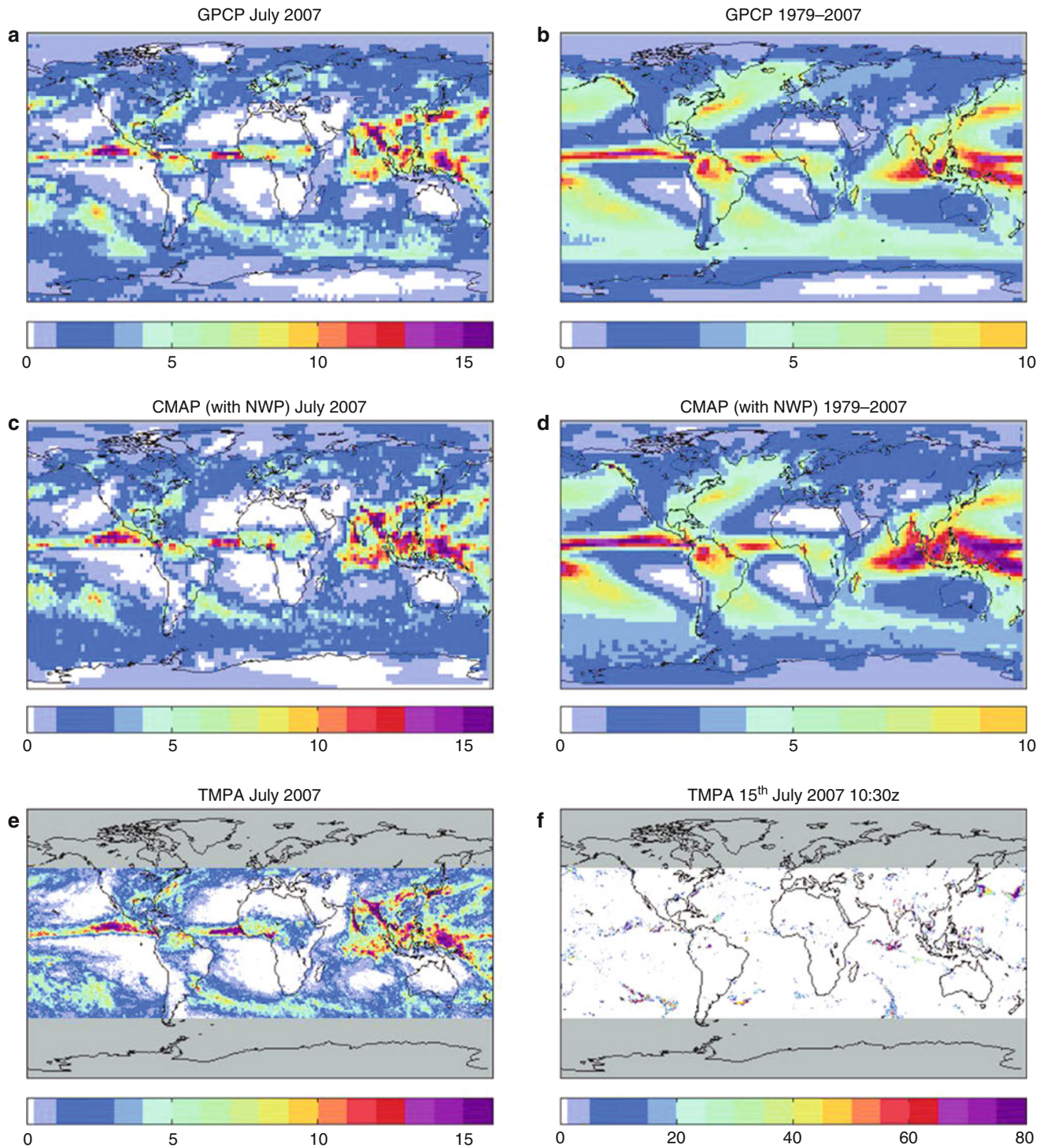
Multisensor datasets

The individual sensor records discussed so far each have limitations which make them unsuitable for use in certain situations. For example, PMW estimates over the ocean might be the more accurate than GEO-IR estimates, but the latter are better suited for studies of the diurnal cycle due to the superior sampling obtained from a GEO satellite. Conveniently, the available remotely sensed estimates of precipitation have different strengths and weaknesses, so that combined datasets can be superior to estimates from individual sensors. Several efforts to intercompare and evaluate various types of precipitation algorithms using remotely sensed information were carried out during the 1990s. The WetNet (Dodge and Goodman, 1994) Precipitation Intercomparison Projects (PIP) evaluated multiple global and near-global precipitation algorithms including merged satellite datasets (Barrett et al., 1994; Kniveton et al., 1994; Smith et al., 1998; Adler et al., 2001). The Global Precipitation Climatology Project (GPCP) similarly sponsored three Algorithm Intercomparison Projects (AIP; Ebert et al., 1996) that compared precipitation estimated from satellite observations against high-resolution observations from rain gauges and radars over limited domains (Arkin and Xie, 1994;

Ebert and Manton, 1998). Among other results, these studies showed that PMW estimates were more accurate than IR estimates on an instantaneous basis, but algorithms which combine PMW and IR estimates were superior. However, these intercomparisons did not show significant differences between individual algorithms of a common type, and it remains the case that several merged satellite products exist without a clear consensus on which is superior, and it is common to see a range of similar datasets used in the literature.

Early combinations: GPCP and CMAP

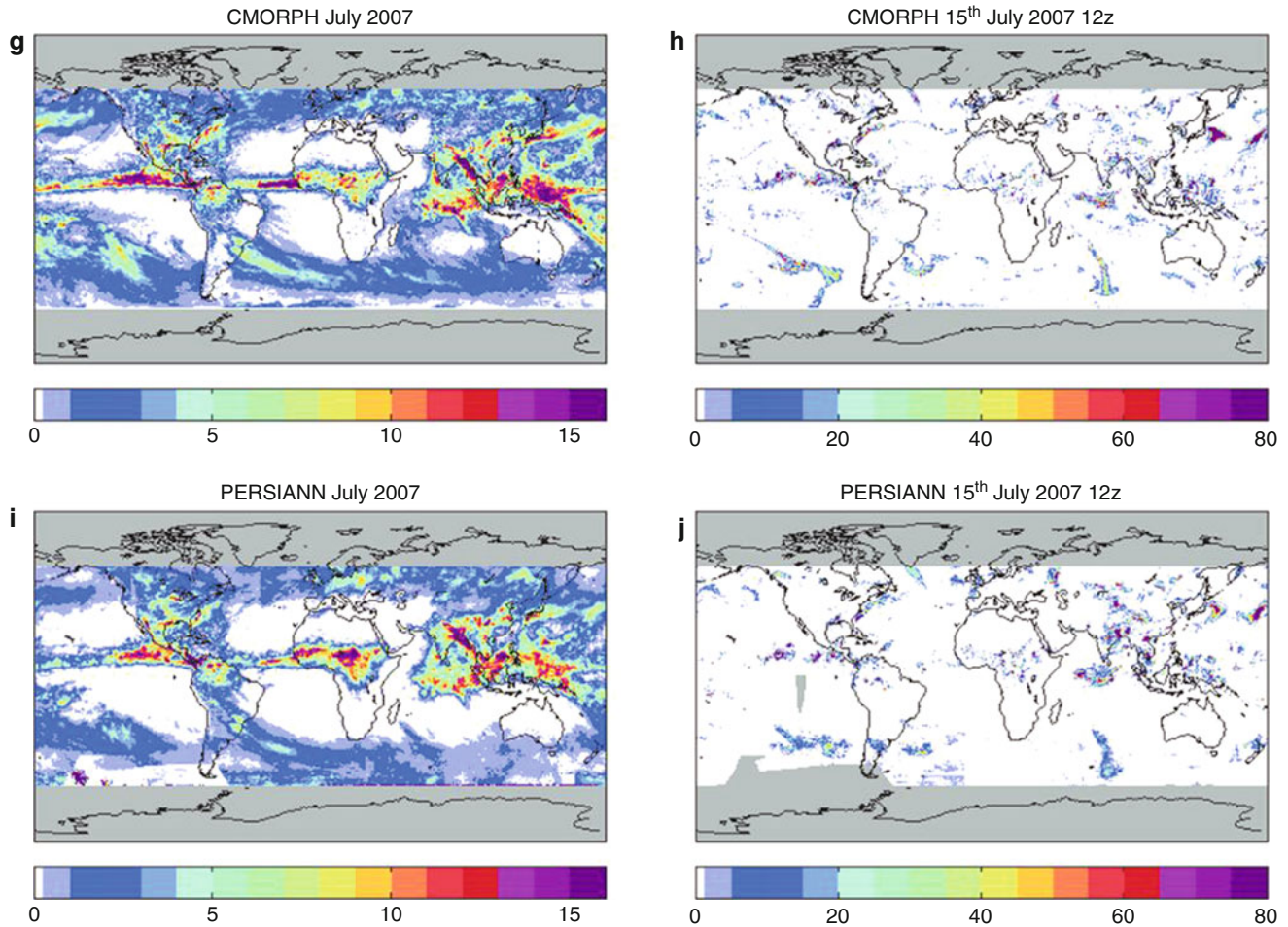
The pre-TRMM (i.e., pre-1997) multisensor combinations of precipitation which included PMW as an input were restricted to combinations of VIS or IR estimates (both geosynchronous and low orbit) and estimates from SSM/I. Many of the early multisource precipitation estimates failed to progress past the research-only phase, but two datasets have endured and are both still in common use: the Global Precipitation Climatology Project (GPCP; Huffman et al., 1997; Adler et al., 2003) and the Climate Prediction Center Merged Analysis of Precipitation (CMAP; Xie and Arkin, 1997b). GPCP and CMAP are both monthly 2.5° resolution and use several common



Rainfall, Figure 6 (Continued)

inputs with somewhat similar merge algorithms and are therefore frequently considered to be similar, although they use different gauge analyses and have different corrections over the tropical oceans. Both datasets start in

1979 and have changing inputs over time as superior data records become available, such as the SSM/I estimates in 1987. The current version of the GPCP 2.5° monthly mean dataset is the version two dataset (Adler et al., 2003)



Rainfall, Figure 6 Maps of precipitation estimates from (a and b) GPCP Version 2, (c and d) CMAP with NCEP/NCAR reanalysis at high latitudes, (e and f) the TMPA, (g and h) CMORPH, and (i and j) PERSIANN. Maps show mean precipitation for (a, c, e, g, and i) July 2007; (b, d) January 1979–December 2007; (f) 10:30–13:30z, 15 July 2007; and (h and j) 12–15z, 15 July 2007. All units are scaled to mm day^{-1} for comparison.

which improved on the first version with a longer record and the addition of TOVS data for improved estimates at mid- and higher latitudes. Both CMAP and GPCP have problems with high-latitude precipitation due to the lack of reliable data: there are few gauges in these sparsely populated regions and available satellite-derived precipitation estimates are of limited use over ice- or snow-covered surfaces. Some studies suggest that model data might be superior to all other estimates at high latitudes (Serreze et al., 2005; Su et al., 2006; Sapiano et al., 2008), and a version of CMAP includes reanalysis precipitation forecasts from the NCAR NCEP reanalysis (Kalnay et al., 1997) over the high latitudes which is the version shown in Figure 6 (CMAP is also available without the reanalysis data). Figure 6 shows the substantial difference in the magnitude of tropical precipitation between GPCP and CMAP as well as the broad differences in precipitation at higher latitudes, although the broad patterns are very similar as would be expected from the use of many common inputs.

Although the monthly, 2.5° products are most commonly used, higher-resolution products also exist as part of the GPCP suite. A pentad (5 days mean) version of GPCP combines similar satellite inputs as the monthly product with a different gauge dataset (Xie et al., 2003). This experimental dataset is also produced on a 2.5° resolution grid and starts in 1979. A one-degree, daily version of the GPCP dataset is also available which starts in 1996 and combines IR from geosynchronous and low orbits with the GPCP Version 2 monthly product and AIRS and TOVS estimates (Huffman et al., 2001). This dataset does not directly use PMW data, although the SSM/I estimates are implicitly used through the inclusion of the monthly GPCP Version 2 product.

High-resolution precipitation products

While GPCP and CMAP are still commonly used for global, monthly, coarse resolution studies (including

validation of numerical model forecasts), recent increases in the availability of PMW data (SSM/I, TMI, AMSU, AMSR) have led to the emergence of several near-global, high-resolution products which have begun to supersede the pentad and one-degree daily products of the GPCP. These products are not global (extending no further than 60°N/S) but have spatial resolutions of at least 0.25° and temporal resolution of at least three hourly and are based on a variety of innovative methods for combining estimates derived from PMW observations with GEO-IR imagery. Generally speaking, these high-resolution precipitation products (HRPPs) use the high spatial and temporal resolution of IR data to resolve deficiencies in resolution of the higher-quality PMW data, although there are substantial differences between the exact methodologies employed. The HRPPs can be categorized into two broad types: adjustment-based techniques where IR data is calibrated using PMW estimates (where the two are often then combined) and motion-based techniques, where the IR data is used to interpolate between successive PMW overpasses.

The Pilot Evaluation of High-Resolution Precipitation Products (PEHRPP) was established to intercompare and validate these datasets. PEHRPP included a number of high-resolution datasets: the TRMM Multisatellite Precipitation Analysis (TMPA; Huffman et al., 2007), the CPC Morphing technique (CMORPH; Joyce et al., 2004), the Hydro-Estimator (Scotfield and Kuligowski, 2003), the NRL-Blended technique (NRL-Blended; Turk and Miller, 2005), Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Hsu et al., 1997; Sorooshian et al., 2000), and the Global Satellite Mapping of Precipitation (GSMaP) project. Studies to compare these datasets are underway, although there is no technique that outperforms the others on a consistent basis (Gottschalck et al., 2005; Brown, 2006; Ebert et al., 2007; Ruane and Roads, 2007; Tian et al., 2007; Sapiiano and Arkin, 2009). Figure 6 shows a single three-hourly and daily accumulation for TMPA (Figure 6f), CMORPH (Figure 6h), and PERSIANN (Figure 6j). These three datasets have different properties; TMPA is a simple combination of PMW and IR starting in 1997 and is gauge adjusted; CMORPH uses IR information to advect PMW information between successive overpasses and has near-global resolution of 8 km, three hourly; PERSIANN uses a neural network (trained with PMW estimates) to translate IR only to rain rates. The different properties of these datasets are clear, with PERSIANN giving somewhat smoother estimates than the other two, which generally agree with each other, although TMPA is closer to GPCP and CMAP over land areas. Despite recent advances there is still much uncertainty in the quality of high-resolution precipitation datasets compared to coarser datasets such as GPCP and CMAP. Reconciliation between the HRPPs and the coarser datasets has still not been achieved, and the algorithm designs employed by GPCP and CMAP remove much of the inherent noise from the inputs, so that they

are considered more reliable than the HRPPs. In short, the reader is advised to use the class of dataset that has closest resolution to their needs and is cautioned against averaging high-resolution data.

Summary

There are various techniques to retrieve rainfall from satellites, each with their own set of attributes that are dictated by the particular needs for the information; for short-term, high spatial resolution applications like flash flood forecasting, the IR methods are generally preferred, while for global, climate scales, the PMW are usually preferred. In this sense, no single approach can be defined as the best one; however, in terms of accuracy on the instantaneous time scale (i.e., at the time the satellite is making its measurement), it is generally accepted that the active MW is the most accurate, followed by passive MW (ocean), passive MW (land), IR, and VIS.

Many of the early methods were developed using sensors that were not necessarily flown for rainfall retrieval but more for tracking cloud features and monitoring atmospheric temperature and moisture. As was described, more recent and near-term missions are being designed specifically for rainfall monitoring and include AMW sensors (e.g., the GPM mission). Additionally, emerging methods such as the blended techniques or multispectral (including PMW and AMW) will yield improvements to the current accuracy of the remote sensing of rainfall and will likely become the standard retrieval method as we enter into the next decade.

Bibliography

- Adler, R. F., and Negri, A. J., 1988. A satellite infrared technique to estimate tropical convective and stratiform rainfall. *Journal of Applied Meteorology*, **27**, 30–51.
- Adler, R. F., Kidd, C., Petty, G., Morrissey, M., and Goodman, H. M., 2001. Intercomparison of global precipitation products: the third precipitation intercomparison project (PIP-3). *Bulletin of the American Meteorological Society*, **82**, 1377–1396.
- Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E., 2003. The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, **4**, 1147–1167.
- Amayenc, P., Diguët, J. P., Marzoug, M., and Tani, T., 1996. A class of single- and dual-frequency algorithms for rain-rate profiling from a spaceborne radar. Part II: tests from airborne radar measurements. *Journal of Atmospheric and Oceanic Technology*, **13**, 142–164.
- Amorati, R., Alberoni, P. P., Levizzani, V., and Nanni, S., 2000. IR-based satellite and radar rainfall estimates of convective storms over northern Italy. *Meteorological Applications*, **7**, 1–18.
- Arkin, P. A., 1979. The relationship between fractional coverage of high cloud and rainfall accumulations during GATE over the B-scale array. *Monthly Weather Review*, **106**, 1153–1171.
- Arkin, P. A., and Janowiak, J., 1991. Analysis of the global distribution of precipitation. *Dynamics of Atmospheres and Oceans*, **16**, 5–16.

- Arkin, P. A., and Meisner, B. N., 1987. The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982–84. *Monthly Weather Review*, **115**, 51–74.
- Arkin, P. A., and Xie, P. P., 1994. The global precipitation climatology project: first algorithm intercomparison project. *Bulletin of the American Meteorological Society*, **75**, 401–419.
- Ba, M. B., and Gruber, A., 2001. GOES multispectral rainfall algorithm (GMSRA). *Journal of Applied Meteorology*, **40**, 1500–1514.
- Barrett, E. C., and Martin, D. W., 1981. *The Use of Satellite Data in Rainfall Monitoring*. London: Academic.
- Barrett, E. C., Dodge, J., Goodman, H. M., Janowiak, J., Kidd, C., and Smith, E. A., 1994. The first WetNet precipitation intercomparison project. *Remote Sensing Reviews*, **11**, 49–60.
- Bauer, P., Moreau, E., and Di Michele, S., 2005. Hydrometeor retrieval accuracy using microwave window and sounding channel observations. *Journal of Applied Meteorology*, **44**, 1016–1032.
- Bellon, A., Lovejoy, S., and Austin, G. L., 1980. Combining satellite and radar data for the short-range forecasting of precipitation. *Monthly Weather Review*, **108**, 1554–1556.
- Brown, J. E. M., 2006. An analysis of the performance of hybrid infrared and microwave satellite precipitation algorithms over India and adjacent regions. *Remote Sensing of Environment*, **101**, 63–81.
- Dodge, J., and Goodman, H. M., 1994. The WetNet project. *Remote Sensing Reviews*, **11**, 5–21.
- Doviak, J. D., and Zrnicek, D. S., 1984. *Doppler Radar and Weather Observations*. New York: Academic.
- Ebert, E. E., and Manton, M. J., 1998. Performance of satellite rainfall estimation algorithms during TOGA COARE. *Journal of the Atmospheric Science*, **55**, 1537–1557.
- Ebert, E. E., Manton, M. J., Arkin, P. A., Allam, R. J., Holpin, C. E., and Gruber, A., 1996. Results from the GPCP algorithm intercomparison programme. *Bulletin of the American Meteorological Society*, **77**, 2875–2887.
- Ebert, E. E., Janowiak, J. E., and Kidd, C., 2007. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, **88**, 47–64.
- Ferraro, R. R., 1997. Special sensor microwave imager derived global rainfall estimates for climatological applications. *Journal of Geophysical Research*, **102**, 16715–16735.
- Ferraro, R. R., and Marks, G. F., 1995. The development of SSM/I rain-rate retrieval algorithms using ground-based radar measurements. *Journal of Atmospheric and Oceanic Technology*, **12**, 755–770.
- Fujita, M., and Satake, M., 1997. Rainfall rate profiling with attenuation-frequency radar using nonlinear LMS technique under a constraint on path-integrated rainfall rate. *International Journal of Remote Sensing*, **18**, 1137–1147.
- Gottschalk, J., Meng, J., Rodell, M., and Houser, P., 2005. Analysis of multiple precipitation products and preliminary assessment of their impact on global land data assimilation system land surface states. *Journal of Hydrometeorology*, **6**, 573–598.
- Griffith, C. G., Woodley, W. L., Grube, P., Martin, D. W., Stout, J., and Sikdar, D. N., 1978. Rain estimation from geosynchronous satellite imagery – visible and infrared studies. *Monthly Weather Review*, **106**, 1153–1171.
- Grody, N. C., 1991. Classification of snow cover and precipitation using the special sensor microwave/imager (SSM/I). *Journal of Geophysical Research*, **96**, 7423–7435.
- Haynes, J. M., and Stephens, G., 2007. Tropical oceanic cloudiness and incidence of precipitation: early results from CloudSat. *Geophysical Research Letters*, **34**, L09811, doi:10.1029/2007GL029335.
- Hitschfeld, W., and Bordan, J., 1954. Errors inherent in the radar measurement of rainfall at attenuating wavelengths. *Journal of Meteorology*, **11**, 58–67.
- Hsu, K., Gao, X., Sorooshian, S., and Gupta, H. V., 1997. Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of Applied Meteorology*, **36**, 1176–1190.
- Huffman, G. J., Adler, R. F., Arkin, P. A., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., and Schneider, U., 1997. The global precipitation climatology project (GPCP) combined precipitation dataset. *Bulletin of the American Meteorological Society*, **78**, 5–20.
- Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B., and Susskind, J., 2001. Global precipitation at one-degree daily resolution from multisatellite observations. *Journal of Hydrometeorology*, **2**, 36–50.
- Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E., and Wolff, D., 2007. The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, **8**, 38–55.
- Iguchi, T., and Meneghini, R., 1994. Intercomparison of single frequency methods for receiving a vertical rain profile from airborne and spaceborne radar data. *Journal of Atmospheric and Oceanic Technology*, **11**, 1507–1516.
- Iguchi, T., Kozu, T., Meneghini, R., Awaka, J., and Okamoto, K., 2000. Rain-profiling algorithm for the TRMM precipitation radar. *Journal of Applied Meteorology*, **39**, 2038–2052.
- Iguchi, T., Oki, R., Smith, E. A., and Furuhashi, Y., 2002. Global precipitation measurement program and the development of dual-frequency precipitation radar. *Journal of the Communication Research Laboratory*, **49**, 37–45.
- Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, **5**, 487–503.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–471.
- Kidder, S. Q., and Vonder Haar, T. H., 1995. *Satellite Meteorology: An Introduction*. New York: Academic.
- Kniveton, D. R., Motta, B. C., Goodman, H. M., Smith, M., and LaFontaine, F. J., 1994. The first WetNet precipitation intercomparison project: generation of results. *Remote Sensing Reviews*, **11**, 243–302.
- Kummerow, C., Olson, W. S., and Giglio, L., 1996. A simplified scheme for obtaining precipitation and vertical hydrometeor profile from passive microwave. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 1213–1232.
- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J., 1998. The tropical rainfall measuring mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, **15**, 809–817.
- Kummerow, C., Hong, Y., Olson, W. S., Yang, S., Adler, R. F., McCollum, J., Ferraro, R., Petty, G., Shin, D., and Wilheit, T. T., 2001. The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. *Journal of Applied Meteorology*, **40**, 1801–1819.
- Kuo, K.-S., Smith, E. A., Haddad, S., Im, E., Iguchi, T., and Mugnai, A., 2004. Mathematical physical framework for retrieval of rain DSD properties from dual-frequency Ku/Ka-band satellite radar. *Journal of the Atmospheric Sciences*, **61**, 2349–2369.

- L'Ecuyer, T. S., and Stephens, G. L., 2002. An estimation-based precipitation retrieval algorithm for attenuating radars. *Journal of Applied Meteorology*, **41**, 272–285.
- Lovejoy, S., and Austin, G. L., 1979. The delineation of rain areas from visible and IR satellite data from GATE and mid-latitudes. *Atmosphere-ocean*, **17**, 77–92.
- Marzoug, M., and Amayenc, P., 1994. A class of single frequency algorithms for rain rate profiling from a spaceborne radar I. Principle and tests from numerical simulations. *Journal of Atmospheric and Oceanic Technology*, **11**, 1480–1506.
- Matrosov, S., 2007. Potential for attenuation-based estimates of rainfall rate from CloudSat. *Geophysical Research Letters*, **34**, L05817, doi:10.1029/2006GL029161.
- McCollum, J. R., and Ferraro, R. R., 2003. The next generation of NOAA/NESDIS SSM/I, TMI and AMSR-E microwave land rainfall algorithms. *Journal of Geophysical Research*, **108**, 8382–8404.
- Meneghini, R., and Kozu, T., 1990. *Spaceborne Weather Radar*. Norwood, MA: Artech House.
- Meneghini, R., Nakamura, K., Ulbrich, C. W., and Atlas, D., 1989. Experimental tests of methods for the measurement of rainfall rate using an airborne dual-wavelength radar. *Journal of Atmospheric and Oceanic Technology*, **6**, 637–651.
- Meneghini, R., Kozu, T., Kumagai, H., and Bonczyk, W. C., 1992. A study of rain estimation methods from space using dual-wavelength radar measurements at near-nadir incidence over ocean. *Journal of Atmospheric and Oceanic Technology*, **9**, 364–382.
- Meneghini, R., Iguchi, T., Kozu, T., Liao, L., Okamoto, K., Jones, J., and Kwiatkowski, J., 2000. Use of the surface reference technique for path attenuation estimates from the TRMM precipitation radar. *Journal of Applied Meteorology*, **39**, 2053–2070.
- Mitrescu, C., Heynes, J., L'ecuyer, T., Turk, J., and Miller, S., 2008. CloudSat precipitation profiling algorithm-model description. In *The 88th American Meteorological Society meeting*, 20–25 Jan 2008, New Orleans.
- Mugnai, A., and Smith, E.A., 1984. Passive microwave radiation transfer in an evolving cloud medium. In *IRS'84: Current Problems in Atmospheric Radiation*. Hampton, VA: A. Deepak Publishing.
- Ohno, Y., Horie, H., Sato, K., Kumagai, H., Kimura, T., Okada, K., Iida, Y., and Kojima, M., 2007. Development of cloud profiling radar for EarthCare. In *Thirty-Third AMS Conference on Radar Meteorology*, 6–10 Aug 2007, Cairns.
- Olson, W. S., 1989. Physical retrieval of rainfall rates over the ocean by multispectral microwave radiometry: application to tropical cyclones. *Journal of Geophysical Research*, **94**, 2267–2280.
- Rosenfeld, D., and Gutman, G., 1994. Retrieving microphysical properties near the tops of potential rain clouds by multispectral analysis of AVHRR data. *Atmospheric Research*, **34**, 259–283.
- Ruane, A. C., and Roads, J. O., 2007. 6-hour to 1-year variance of five global precipitation sets. *Earth Interactions Journal*, **11**, 1.
- Sapiano, M. R. P., and Arkin, P. A., 2009. An inter-comparison and validation of high resolution satellite precipitation estimates with three-hourly gauge data. *Journal of Hydrometeorology*, **10**(1), 149–166.
- Sapiano, M. R. P., Smith, T. M., and Arkin, P. A., 2008. A new merged analysis of precipitation utilizing satellite and reanalysis data. *Journal of Geophysical Research*, **113**, D22103, doi:10.1029/2008JD010310.
- Scofield, R. A., 1987. The NESDIS operational convective precipitation technique. *Monthly Weather Review*, **115**, 1773–1792.
- Scofield, R. A., and Kuligowski, R. J., 2003. Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. *Weather and Forecasting*, **18**, 1037–1051.
- Serreze, M. C., Barrett, A., and Lo, F., 2005. Northern high latitude precipitation as depicted by atmospheric reanalyses and satellite retrievals. *Monthly Weather Review*, **133**, 3407–3430.
- Smith, E. A., Xiang, X., Mugnai, A., and Tripoli, G. J., 1994. Design of an inversion-based precipitation profile retrieval algorithm using an explicit cloud model for initial guess microphysics. *Meteorology and Atmospheric Physics*, **54**, 53–78.
- Smith, E. A., et al., 1998. Results of WetNet PIP-2 project. *Journal of the Atmospheric Sciences*, **55**, 148–1536.
- Sorooshian, S., Hsu, K., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, **81**, 2035–2046.
- Spencer, R. W., Martin, D. W., Hinton, B., and Weinman, J. A., 1983. Satellite microwave radiances correlated with radar rain rates over land. *Nature*, **304**, 141–143.
- Spencer, R. W., Goodman, H. M., and Hood, R. E., 1989. Precipitation retrieval over land and ocean with the SSM/I: identification and characteristics of the scattering signal. *Journal of Atmospheric and Oceanic Technology*, **6**, 254–273.
- Stephens, G., et al., 2002. The CloudSat mission and the A-Train. *Bulletin of the American Meteorological Society*, **83**, 1771–1790.
- Su, F., Adam, J., Trenberth, K., and Lettenmaier, D., 2006. Evaluation of surface water fluxes of the pan-Arctic land region with a land surface model and ERA-40 reanalysis. *Journal of Geophysical Research*, **111**, D05110, doi:10.1029/2005JD006387.
- Tian, Y., Peters-Lidard, C., Choudhury, B., and Garcia, M., 2007. Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications. *Journal of Hydrometeorology*, **8**, 1165–1183.
- Todd, M. C., Barrett, E. C., Beaumont, M. J., and Green, J. L., 1995. Satellite identification of rain days over the upper Nile river basin using an optimum infrared rain/no-rain threshold temperature model. *Journal of Applied Meteorology*, **34**, 2600–2611.
- Todd, M. C., Barrett, E. C., Beaumont, M. J., and Bellerby, T. J., 1999. Estimation of daily rainfall over the upper Nile river basin using a continuously calibrated satellite infrared technique. *Meteorological Applications*, **6**, 201–210.
- Turk, F. J., and Miller, S. D., 2005. Toward improved characterization of remotely sensed precipitation regimes with MODIS/AMSR-E blended data techniques. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 1059–1069.
- Vicente, G. A., Scofield, R. A., and Menzel, W. P., 1998. The operational GOES infrared rainfall estimation technique. *Bulletin of the American Meteorological Society*, **79**, 1883–1898.
- Vicente, G. A., Davenport, J. C., and Scofield, R. A., 2002. The role of orographic and parallax correction on real-time high resolution satellite rainfall rate distribution. *International Journal of Remote Sensing*, **23**, 221–230.
- Vila, D., Ferraro, R. R., and Joyce, R., 2007. Evaluation and improvement of AMSU precipitation retrievals. *Journal of Geophysical Research*, **112**, D20119, doi:10.1029/2007JD008617.
- Weinman, J., and Guetter, A., 1977. Determination of rainfall distributions from microwave radiation measured by the Nimbus 6 ESMR. *Journal of Applied Meteorology*, **16**, 437–442.
- Weng, F., and Grody, N. C., 1994. Retrieval of cloud liquid water using the special sensor microwave imager (SSM/I) data. *Journal of Geophysical Research*, **99**, 25535–25551.
- Weng, F., Zhao, L., Poe, G., Ferraro, R., Li, X., and Grody, N., 2003. AMSU cloud and precipitation algorithms. *Radio Science*, **338**, 8068–8079.
- Wilheit, T. T., Chang, A., Rao, M., Rodgers, E., and Theon, J., 1977. A satellite technique for quantitatively mapping rainfall rates over the oceans. *Journal of Applied Meteorology*, **16**, 551–560.
- Wilheit, T. T., Kummerow, C., and Ferraro, R., 2003. Rainfall algorithms for AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 204–214.

- Woodley, W. L., and Sancho, R., 1971. A first step towards rainfall estimation from satellite cloud photographs. *Weather*, **26**, 279–289.
- Xie, P., and Arkin, P. A., 1997a. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bulletin of the American Meteorological Society*, **78**, 2539–2558.
- Xie, P., and Arkin, P. A., 1997b. Global pentad precipitation analysis based on gauge observations, satellite estimates and model outputs. In *Extended Abstracts, American Geophysical Union 1997 Fall Meeting*, 8–12 Dec 1997, San Francisco, CA: AGU.
- Xie, P., Janowiak, J. E., Arkin, P. A., Adler, R., Gruber, A., Ferraro, R., Huffman, G., and Curtis, S., 2003. GPCP pentad precipitation analyses: an experimental dataset based on gauge observations and satellite estimates. *Journal of Climate*, **16**, 2197–2214.
- Zhao, L., and Weng, F., 2002. Retrieval of ice cloud parameters using the AMSU. *Journal of Applied Meteorology*, **41**, 384–395.

Cross-references

Climate Monitoring and Prediction
Global Earth Observation System of Systems (GEOSS)
Land Surface Emissivity
Microwave Radiometers
Observational Systems, Satellite
Optical/Infrared, Radiative Transfer
Water Resources

RANGELANDS AND GRAZING

Hunt E. Raymond, Jr.
USDA-ARS Hydrology and Remote Sensing
Laboratory, Beltsville, MD, USA

Synonyms

Grasslands; Grazing lands; Pampas; Prairie; Savanna; Shrublands; Steppes

Definition

Rangelands. Type of land cover dominated by grasses, grasslike plants, broadleaf herbaceous plants (forbs), and shrubs, where the land is managed as a natural ecosystem for multiple uses including wildlife habitat, biodiversity, recreation, and grazing by livestock.

Pasturelands. Type of land cover dominated by grasses, grasslike plants, and broadleaf herbaceous plants, where the land is managed as an agricultural system for livestock production.

Plant community. Co-occurring plant species and their relative abundance and which is usually recognized by the dominant plant species or functional type.

Ecological site. A combination of soil, climate, and hydrological factors that have the potential for a distinctive climax plant community with a given amount of net primary production.

Rangeland health. The degree to which the integrity of the soil is maintained and ecological processes are sustained.

Introduction

Rangelands encompass many different land cover types throughout the world, including grasslands, savannas, shrublands, tundras, marshes, and meadows (Holechek et al., 2004). Based upon the NASA MODIS land cover product, rangelands cover about 48 % of the Earth's land surface, not counting Antarctica (Friedl et al., 2010).

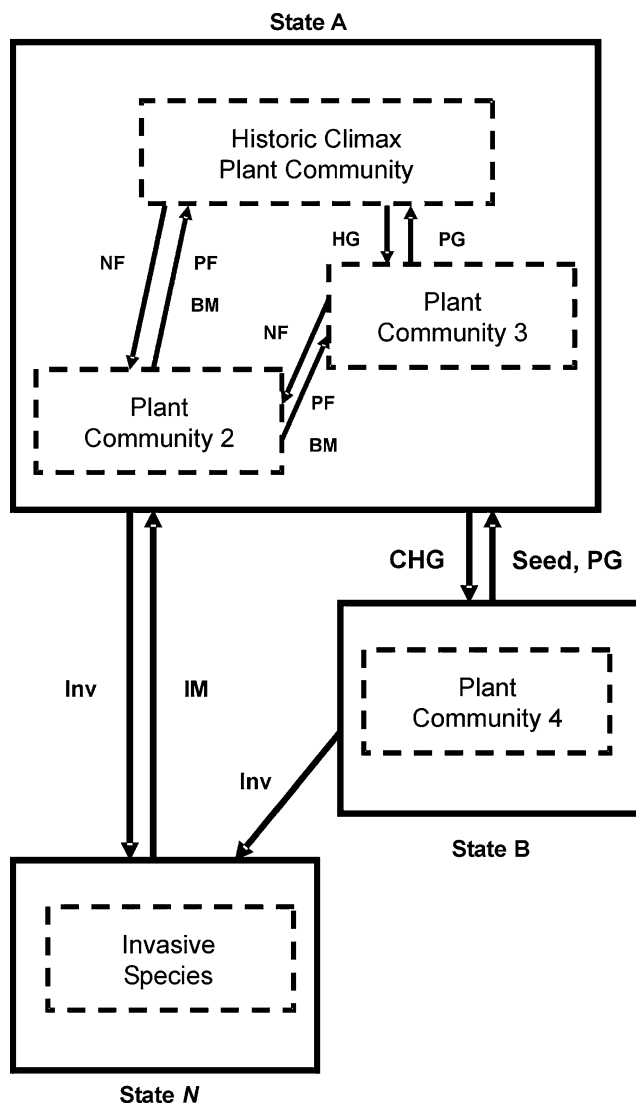
Most rangeland cover types are found in regions where the annual precipitation is low and the interannual variability of precipitation is high, which makes these regions unsuitable for forestry or agriculture. Livestock grazing is a land use that provides a sustainable means of producing food and fiber in rangeland cover types. Confusingly, many nongovernmental organizations use the term rangelands to denote a land-use category of livestock grazing (Lund, 2007).

Inventory, assessment, and management of grazing on rangelands within the USA are based on ecological sites, which are generally recognized and delimited by the historic climax plant community (NRCS, 2003). Another paradigm for management is rangeland health, which is based on prevention of soil erosion and the sustainability of ecosystem processes such as energy flow (primary production), the hydrologic cycle, and biogeochemical cycles (BLM, 2005; NRC, 1994). The concepts of rangeland health and ecological sites were developed in parallel in response to the shortcomings of "Rangeland Condition and Trend" (Briske et al., 2005).

Ecological sites

Management of rangelands was based initially on the theory of plant succession from Clements, (1916). Starting from bare substrate (primary succession) or bare soil (secondary succession), the dominant species change over time from short-lived to long-lived plants, ultimately reaching a climax plant community determined by climate. Climax plant communities were assumed to have the highest sustainable productivity, the greatest resistance to invasive species and pests, and the best protection against soil erosion; thus, rangeland management was based on maintaining or restoring the climax plant community. However, it was recognized that natural disturbances such as drought, grazing, and fire had to be incorporated into the definition of the climax community (NRCS, 2003; Briske et al., 2005). Ecological sites are identified and delimited by (1) the dominant plant community at the onset of European immigration and settlement (NRCS, 2003) and (2) soils and climate delineated by major land resource areas (NRCS, 2006).

An ecological state is comprised of group of plant communities which are in a dynamic equilibrium with the historical amounts and types of disturbances at a given ecological site. If disturbances are substantially greater than historical levels, there may be a transition to a different ecological state, with new groups of plant communities (NRCS, 2003; Briske et al., 2005).



Rangelands and Grazing, Figure 1 State-and-transition diagram for an ecological site. The large boxes with solid lines are stable ecological states (A, B, ..., N) within which dynamic changes in plant communities result from natural disturbances. Changes from one state to another (transition) are usually irreversible without large management inputs. Boxes with dashed lines are used to represent a specific plant community that occurs within a stable ecological state. The historical climax plant community is the ideal management goal for sustainable use. Arrows show possible changes and the symbols represent a mechanism for the possible change. Symbols are *NF* no fire, *PF* prescribed fire, *BM* brush management, *HG* heavy grazing, *PG* prescribed grazing, *CHG* continuous heavy grazing, *Seed* reseeding, *Inv* invasion, and *IM* integrated management.

The relationships among plant communities and ecological states for an ecological site are depicted in a state-and-transition model (Figure 1). In Figure 1, State A includes the various plant communities found at an ecological site with historical levels of disturbance, including the historic climax plant community. Within State A, plant community 2 could be the result of woody shrub establishment allowed by reduced fire frequency and plant community 3 could be an earlier stage of succession after disturbance. However, continuous heavy grazing may result in soil erosion, so the site crosses a threshold into a new State B with plant community 4;

changes of grazing intensity alone are unlikely to restore the site to State A without reseeding (Figure 1). Alternatively, the ecological site could be invaded by a nonnative weed species (State N), requiring integrated management (a combination of biological, cultural, and chemical control) for restoration (Figure 1).

Because an ecological site is defined by the potential to have a climax plant community, and because an ecological state may have many different plant communities, rarely will land cover classification from satellite remote sensing have a major role for mapping either ecological sites or states (Hunt et al., 2003; Pickup et al., 1994).

A combination of imagery, geospatial databases of soils and climate, and computer simulation modeling could be used to determine potential vegetation for ecological sites (Jensen et al., 2001). Simple detection of land cover change between two dates is not useful for detecting a state transition because of year-to-year variability; analysis of long-term imagery is important to distinguish a state transition (Washington-Allen et al., 2006).

Rangeland health

Rangeland health is monitored by relative rankings of 17 indicators based on (1) soils and soil structure; (2) hydrological, nutrient, and carbon cycling (including net primary production); and (3) functional plant diversity and presence of invasive species (BLM, 2005). Of all of the indicators, ground cover fractions have been demonstrated as having the greatest utility for economical sampling and objective measurement on a landscape scale (Booth and Tueller, 2003; Pickup et al., 1994). Other aspects of rangeland health could be monitored operationally with the appropriate sensors, such as changes in vegetation productivity, invasion of noxious weeds, and cover of senesced and living plant material (Hunt et al., 2003; Palmer and Fortesque, 2004). If it is accepted that unbiased, economical monitoring rangelands is required, and remote sensing provides these methods, then the framework for rangeland assessment needs to be reformulated based on what remote sensing can accomplish (Pickup et al., 1994, p. 499).

Rangeland productivity

Primary production of rangelands can be estimated using the amount of absorbed photosynthetically active radiation integrated over time, which is the basis of MODIS primary production (MOD17A3) data product (Reeves et al., 2001, 2006; Running et al., 2004). NDVI is approximately equal to the fraction of incident photosynthetically active radiation that is absorbed by the canopy, so NDVI integrated over time is also used to estimate primary production (Paruelo et al., 1997; Piñeiro et al., 2006; Tieszen et al., 1997).

Net primary production provides the upper limit on available forage on a sustainable basis and hence determines the stocking rate or carrying capacity. Determination of standing biomass by remote sensing is possible, but grazing based on standing biomass may not be sustainable. A 455 kg cow (one animal unit) consumes about 9.1 kg of forage per day (Holechek et al., 2004); 273 kg of forage is one animal unit month (AUM). The above-ground net primary production divided by 273 kg/month sets the stocking rate in AUM per area; factors such as slope and distance to water reduce the stocking rate from the maximum (Holechek et al., 2004). Many of these factors can be estimated from geospatial databases allowing determination of recommended stocking rates (Hunt et al., 2003; Hunt and Miyake, 2006). Environmental data records of NDVI from AVHRR, MODIS, SPOT

Vegetation, and VIIRS (Tucker et al., 2005; Brown et al., 2006) are important because the long time series available can assess variability of net primary production due to variability in rainfall.

Detection of invasive weeds

Invasive shrubs and weeds dominate the vegetation of many rangeland plant communities, affecting native biodiversity and reducing the amount of available forage (DiTomaso, 2000). Some invasive plants have spectrally distinctive flowers or leaves that can be detected with imaging spectroscopy (Andrew and Ustin, 2006; Everitt et al., 2002; Hunt et al., 2004; Lass et al., 2002; Parker Williams and Hunt, 2002; Underwood et al., 2003). Imaging spectrometers cover small areas and require extensive expertise for image processing; thus, they may not be useful for operational maps of detectable invasive species (Hunt et al., 2007). However, distribution maps from small areas may be used to test and refine geospatial models for the prediction of sites susceptible to invasion (Bradley and Mustard, 2006; Hunt et al., 2010). Other species such as cheatgrass (Bradley and Mustard, 2006) and salt cedar (Everitt et al., 2002; Groeneveld and Watson, 2008) can be detected using differences in phenology compared to the co-occurring native vegetation.

Ecosystem degradation

The cover fractions of bare soil, senesced plant matter (plant litter or residue), and green vegetation are key indicators of rangeland health (Booth and Tueller, 2003). Bare soil is very susceptible to erosion, and erosion increases the amount of bare soil visible from aircraft and satellites (Vrieling et al., 2007). Green vegetation cover is usually measured by scaling NDVI to account for soil background reflectance (Jiang et al., 2006; Montandon and Small, 2008; Pickup et al., 1998). Senesced plant matter is spectrally similar to bare soil at visible and near-infrared wavelengths; narrowband reflectances in the shortwave infrared may be used to detect senesced plant matter based on the absorption feature of cellulose at a wavelength of 2.1 μm (Daughtry et al., 2004; Nagler et al., 2000).

Spectral unmixing of remotely sensed imagery allows the cover fractions of bare soil, senesced plant matter, and green vegetation to be determined simultaneously (Asner and Heidebrecht, 2002; Roberts et al., 1993). Furthermore, there are a large number of studies that have successfully unmixed multispectral data based on the phenology of vegetation (Arsenault and Bonn, 2005; de Asis and Omasa, 2007; Kuemmerle et al., 2006; Metternicht and Fermont, 1998; Numata et al., 2007).

Very large-scale aerial photography

Very large-scale aerial (VLSA) photographs have very large map scales from 1:50 to 1:500 (pixel sizes of 1–50 mm) and are an important method of rangeland monitoring for either ecological-state or rangeland-health paradigms. Whereas aerial photography has been used

for over 50 years, digital cameras may reduce the cost for image acquisition, processing, and storage. At very large scales, composition of spectrally similar plant species may be identified (Petersen et al., 2005). Individual plants of an invasive species may be detected at the initial stages of establishment (Blumenthal et al., 2007; Naylor et al., 2005; Booth et al., 2010).

Landscape digital photographs may be compared quantitatively with digitized historic landscape photographs for change detection (Clark and Hardegree, 2005). With special robotic camera mount, landscape panoramas may be acquired at full magnification, resulting in a seamless image for analysis (Nichols et al., 2009). Analysis of large numbers of images may be very labor intensive without automation; supervised image classification is usually problematic because only three bands are acquired. "SamplePoint" software (Booth et al., 2006) reduces the time required to estimate cover from VLSA photographs.

Two types of airborne platforms for VLSA photography are being evaluated: unmanned aerial systems (Hardin and Jackson, 2005; Rango et al., 2006) and light manned aircraft (Booth et al., 2006, 2008). Light aircraft and unmanned aerial systems may cover transects or provide wall-to-wall coverage (Laliberte et al., 2010). Ground cover fractions from VLSA photography are highly correlated to ground measurements (Booth et al., 2008). However, because the area covered by each photograph is small, ground cover is not strongly correlated with satellite vegetation indices (Sivanpillai and Booth, 2008).

Management of economic risk with AVHRR NDVI

Crop insurance is used as a risk management tool in which payments are based on actual crop losses. In 2007, USDA Risk Management Agency started a pilot insurance program for pasture, rangeland, and forages (RMA, 2010) in order to manage risks from drought. In nine US states, rangeland managers may choose whether to estimate drought from either a rainfall index or a vegetation index (RMA, 2010). Then, the managers select one or more 3 months time periods when rainfall is important for pasture, rangeland, and forage production in their operations.

Grasses and forbs grow when water is available in the soil, so AVHRR NDVI is an indicator of rainfall in rangelands and low NDVI is an indicator of drought (Davenport and Nicholson, 1993; Di et al., 1994; Ji and Peters, 2003; Wang et al., 2003). Currently, 14 days composites of AVHRR NDVI are made and aggregated to an 8 km by 8 km grid cell by the USGS Earth Resources Observation and Science (EROS) Center (Sioux Falls, South Dakota). Based on the historical averages for a grid cell, a trigger grid index is determined. If the final value of NDVI falls below the trigger grid index, payments to the consumer are made based on area insured. Actual production losses are not used to calculate payments (RMA, 2010).

Conclusion

Primary production and pasture, range, and forage insurance are based on the frequent coverage provided by AVHRR, MODIS, SPOT Vegetation, and VIIRS. In the future, economic losses from drought may be better determined from primary production estimates using these sensors, to better estimate the forage available for livestock and wildlife. Rangeland management is currently based on ecological sites so very large-scale photography is the best remote sensing method; but it is difficult to cover large areas. Rangeland health needs to be better integrated with rangeland management based on ecological sites (Herrick et al., 2006). With the large area of rangelands globally, monitoring rangeland health will be unbiased and cost-effective using remote sensing, but only if the cost of the data is reasonable (Palmer and Fortesque, 2004).

Bibliography

- Andrew, M. E., and Ustin, S. L., 2006. Spectral and physiological uniqueness of perennial pepperweed (*Lepidium latifolium*). *Weed Science*, **54**, 1051–1062.
- Arsenault, E., and Bonn, F., 2005. Evaluation of soil erosion protective cover by crop residues using vegetation indices and spectral mixture analysis of multispectral and hyperspectral data. *Catena*, **62**, 157–172.
- Asner, G. P., and Heidebrecht, K. B., 2002. Spectral unmixing of vegetation, soil and dry carbon cover in arid regions: comparing multispectral and hyperspectral observations. *International Journal of Remote Sensing*, **23**, 3939–3958.
- BLM, 2005. *Interpreting Indicators of Rangeland Health, Version 4. Technical Reference 1734–6. USDI Bureau of Land Management, National Science and Technology Center*. Denver, CO: United States Department of Interior.
- Blumenthal, D., Booth, D. T., Cox, S. E., and Ferrier, C. E., 2007. Large-scale aerial images capture details of invasive plant populations. *Rangeland Ecology & Management*, **60**, 523–528.
- Booth, D. T., and Cox, S. E., 2006. Very large scale aerial photography for rangeland monitoring. *Geocarto International*, **21**, 27–34.
- Booth, D. T., and Cox, S. E., 2008. Image-based monitoring to measure ecological change in rangeland. *Frontiers in Ecology and the Environment*, **6**, 185–190.
- Booth, D. T., and Tueller, P. T., 2003. Rangeland monitoring using remote sensing. *Arid Land Research and Management*, **17**, 455–467.
- Booth, D. T., Cox, S. E., and Berryman, R. D., 2006. Point sampling digital imagery using "SamplePoint". *Environmental Monitoring and Assessment*, **123**, 97–108.
- Booth, D. T., Cox, S. E., Meikle, T., and Zuuring, H. R., 2008. Ground-cover measurements: assessing correlation among aerial and ground-based methods. *Environmental Management*, **42**, 1091–1100.
- Booth, D. T., Cox, S. E., and Teel, D., 2010. Aerial assessment of leafy spurge (*Euphorbia esula* L.) on Idaho's deep fire burn. *Native Plants Journal*, **11**, 327–338.
- Bradley, B. A., and Mustard, J. F., 2006. Characterizing the landscape dynamics of an invasive plant and risk of invasion using remote sensing. *Ecological Applications*, **16**, 1132–1147.
- Briske, D. D., Fuhlendorf, S. D., and Smeins, F. E., 2005. State-and-transition models, thresholds, and rangeland health: a synthesis of ecological concepts and perspectives. *Rangeland Ecology & Management*, **58**, 1–10.

- Brown, M. E., Pinzón, J. E., Didan, K., Morisette, J. T., and Tucker, C. J., 2006. Evaluation of the consistency of long-term NDVI time series derived from AVHRR, SPOT-vegetation, SeaWiFS, MODIS, and Landsat ETM+ sensors. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1787–1793.
- Clark, P. E., and Hardegrave, S. P., 2005. Quantifying vegetation change by point sampling landscape photography time series. *Rangeland Ecology & Management*, **58**, 588–597.
- Clements, F. E., 1916. *Plant Succession: An Analysis of the Development of Vegetation*. Washington, DC: Carnegie Institute of Washington. Carnegie Institute Publication, Vol. 242.
- Daughtry, C. S. T., Hunt, E. R., Jr., and McMurtrey, J. E., III, 2004. Assessing crop residue cover using shortwave infrared reflectance. *Remote Sensing of Environment*, **90**, 126–134.
- Davenport, M. L., and Nicholson, S. E., 1993. On the relation between rainfall and the normalized difference vegetation index for diverse vegetation types in east Africa. *International Journal of Remote Sensing*, **14**, 2369–2389.
- de Asis, A. M., and Omasa, K., 2007. Estimation of vegetation parameter for modeling soil erosion using linear spectra mixture analysis of Landsat ETM data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **62**, 309–324.
- Di, L., Runquist, D. C., and Han, L., 1994. Modelling relationships between NDVI and precipitation during vegetation growth cycles. *International Journal of Remote Sensing*, **15**, 2121–2136.
- DiTomaso, J. M., 2000. Invasive weeds in rangelands: species, impacts and management. *Weed Science*, **48**, 255–265.
- Everitt, J. H., Yang, C., Escobar, D. E., and Davis, M. R., 2002. Using remote sensing to detect and map invasive plant species. *Annals of the Arid Zone*, **41**, 32–342.
- Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X., 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, **114**, 168–182.
- Groeneveld, D. P., and Watson, R. P., 2008. Near-infrared discrimination of leafless saltcedar in wintertime Landsat TM. *International Journal of Remote Sensing*, **29**, 3577–3588.
- Hardin, P. J., and Jackson, M. W., 2005. An unmanned aerial vehicle for rangeland photography. *Rangeland Ecology & Management*, **58**, 439–442.
- Herrick, J. E., Bestelmeyer, B. T., Archer, S., Tugel, A. J., and Brown, J. R., 2006. An integrated framework for science-based arid land management. *Journal of Arid Environments*, **65**, 319–335.
- Holechek, J. L., Piper, R. D., and Herbel, C. H., 2004. *Range Management: Principles and Practices*, 5th edn. Upper Saddle River, NJ: Prentice Hall.
- Hunt, E. R., Jr., and Miyake, B. A., 2006. Comparison of stocking rates from remote sensing and geospatial data. *Rangeland Ecology & Management*, **59**, 11–18.
- Hunt, E. R., Jr., Everitt, J. H., Ritchie, J. C., Moran, M. S., Booth, D. T., Anderson, G. L., Clark, P. E., and Seyfried, M. S., 2003. Applications and research using remote sensing for rangeland management. *Photogrammetric Engineering & Remote Sensing*, **69**, 675–693.
- Hunt, E. R., Jr., McMurtrey, J. E., III, Parker Williams, A. E., and Corp, L. A., 2004. Spectral characteristics of leafy spurge (*Euphorbia esula*) leaves and flower bracts. *Weed Science*, **52**, 492–497.
- Hunt, E. R., Jr., Daughtry, C. S. T., Kim, M. S., and Parker Williams, A. E., 2007. Using canopy reflectance models and spectral angels to assess potential of remote sensing to detect invasive weeds. *Journal of Applied Remote Sensing*, **1**, 013506.
- Hunt, E. R., Jr., Gillham, J. H., and Daughtry, C. S. T., 2010. Improving potential geographic distribution models for invasive plants by remote sensing. *Rangeland Ecology & Management*, **63**, 505–513.
- Jensen, M. E., DiBenedetto, J. P., Barber, J. A., Montagne, C., and Bourgeron, P. S., 2001. Spatial modeling of rangeland potential vegetation environments. *Journal of Range Management*, **54**, 528–536.
- Ji, L., and Peters, A. J., 2003. Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices. *Remote Sensing of Environment*, **87**, 85–98.
- Jiang, Z., Huete, A. R., Chen, J., Chen, Y., Li, J., Yan, G., and Zhang, X., 2006. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sensing of Environment*, **101**, 366–378.
- Kuemmerle, T., Röder, A., and Hill, J., 2006. Separating grassland and shrub vegetation by multivariate pixel-adaptive spectral mixture analysis. *International Journal of Remote Sensing*, **27**, 3251–3271.
- Laliberte, A. S., Herrick, J. E., and Rango, A., 2010. Acquisition, orthorectification, and classification of unmanned aerial vehicle (UAV) imagery for rangeland monitoring. *Photogrammetric Engineering & Remote Sensing*, **76**, 661–672.
- Lass, L. W., Thill, D. C., Shafii, B., and Prather, T. S., 2002. Detecting spotted knapweed (*Centaurea maculosa*) with hyperspectral remote sensing technology. *Weed Technology*, **16**, 426–432.
- Lund, H. G., 2007. Accounting for the world's rangelands. *Rangelands*, **29**(1), 3–10.
- Metternicht, G. I., and Fermont, A., 1998. Estimating erosion surface features by linear mixture modeling. *Remote Sensing of Environment*, **64**, 254–265.
- Montandon, L. M., and Small, E. E., 2008. The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI. *Remote Sensing of Environment*, **112**, 1835–1845.
- Nagler, P. L., Daughtry, C. S. T., and Goward, S. N., 2000. Plant litter and soil reflectance. *Remote Sensing of Environment*, **71**, 207–215.
- National Research Council (NRC), 1994. *Rangeland Health, New Methods to Classify, Inventory, and Monitor Rangelands*. Washington, DC: National Academy Press.
- Naylor, B. J., Endress, B. A., and Parks, C. G., 2005. Multiscale detection of sulfur cinquefoil using aerial photography. *Rangeland Ecology & Management*, **58**, 447–451.
- Nichols, M. H., Ruyle, G. B., and Nourbakhsh, I. R., 2009. Very-high-resolution panoramic photography to improve conventional rangeland monitoring. *Rangeland Ecology & Management*, **62**, 579–582.
- NRCS, 2003. *National Range and Pasture Handbook*, Revision 1. USDA National Resource Conservation Service, Grazing Lands Technology Institute. Washington, DC: United States Department of Agriculture.
- NRCS, 2006. *Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin*. USDA Agriculture Handbook, USDA Natural Resource Conservation Service. Washington, DC: United States Department of Agriculture, Vol. 296.
- Numata, I., Roberts, D. A., Chadwick, O. A., Schimel, J., Sampaio, F. R., Leonidas, F. C., and Soares, J., 2007. Characterization of pasture biophysical properties and the impact of grazing intensity using remotely sensed data. *Remote Sensing of Environment*, **109**, 314–327.
- Palmer, A. R., and Fortesque, A., 2004. Remote sensing and change detection in rangelands. *African Journal of Range & Forage Science*, **21**, 123–128.
- Parker Williams, A., and Hunt, E. R., Jr., 2002. Estimation of leafy spurge cover from hyperspectral imagery using mixture tuned matched filtering. *Remote Sensing of Environment*, **82**, 446–456.

- Paruelo, J. M., Epstein, H. E., Lauenroth, W. K., and Burke, I. C., 1997. ANPP estimates from NDVI for the central grassland region of the United States. *Ecology*, **78**, 953–958.
- Petersen, S. L., Stringham, T. K., and Laliberte, A. S., 2005. Classification of willow species using large-scale aerial photography. *Rangeland Ecology & Management*, **58**, 582–587.
- Pickup, G., Bastin, G. N., and Chewings, V. H., 1994. Remote-sensing-based condition assessment for nonequilibrium rangelands under large-scale commercial grazing. *Ecological Applications*, **4**, 497–517.
- Pickup, G., Bastin, G. N., and Chewings, V. H., 1998. Identifying trends in land degradation in non-equilibrium rangelands. *Journal of Applied Ecology*, **35**, 365–377.
- Pineiro, G., Oesterheld, M., and Paruelo, J. M., 2006. Seasonal variation in aboveground production and radiation-use efficiency of temperate rangelands estimated through remote sensing. *Ecosystems*, **9**, 357–373.
- Rango, A., Laliberte, A., Steele, C., Herrick, J. E., Bestelmeyer, B., Schmutge, T., Roanhorse, A., and Jenkins, V., 2006. Using unmanned aerial vehicles for rangelands: current applications and future potentials. *Environmental Practice*, **8**, 159–168.
- Reeves, M. C., Winslow, J. C., and Running, S. W., 2001. Mapping weekly rangeland vegetation productivity using MODIS algorithms. *Journal of Range Management*, **54**, A90–A105.
- Reeves, M. C., Zhao, M., and Running, S. W., 2006. Applying improved estimates of MODIS productivity to characterize grassland vegetation dynamics. *Rangeland Ecology & Management*, **59**, 1–10.
- RMA, 2010. *Vegetation Index Insurance Standards Handbook 2011 and Succeeding Crop Years*. Washington, DC: United States Department of Agriculture. FCIC-18140 (06–2010) Federal Crop Insurance Corporation, Risk Management Agency.
- Roberts, D. A., Smith, M. O., and Adams, J. B., 1993. Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data. *Remote Sensing of Environment*, **44**, 255–269.
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H., 2004. A continuous satellite-derived measure of global terrestrial primary productivity. *Bioscience*, **54**, 547–560.
- Sivanpillai, R., and Booth, D. T., 2008. Characterizing rangeland vegetation using Landsat and 1-mm VLSA data in central Wyoming (USA). *Agroforestry Systems*, **73**, 55–64.
- Tieszen, L. L., Reed, B. L., Bliss, N. B., Wylie, B. K., and DeJong, D. D., 1997. NDVI, C3 and C4 production, and distributions in Great Plains grassland land cover classes. *Ecological Applications*, **7**, 59–78.
- Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., Vermote, E. F., and Saleous, N. E., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, **26**, 4485–4498.
- Underwood, E., Ustin, S., and DiPietro, D., 2003. Mapping nonnative plants using hyperspectral imagery. *Remote Sensing of Environment*, **86**, 150–161.
- Vrieling, A., Rodrigues, S. C., Bartholomeus, H., and Sterk, G., 2007. Automatic identification of erosion gullies with ASTER imagery in the Brazilian Cerrados. *International Journal of Remote Sensing*, **28**, 2723–2738.
- Wang, J., Rich, P. M., and Price, K. P., 2003. Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA. *International Journal of Remote Sensing*, **24**, 2345–2364.
- Washington-Allen, R. A., West, N. E., Ramsey, R. D., and Efroymson, R. A., 2006. A protocol for retrospective remote sensing-based ecological monitoring of rangelands. *Rangeland Ecology & Management*, **59**, 19–29.

Cross-references

[Climate Data Records](#)
[Soil Properties](#)
[Vegetation Indices](#)
[Vegetation Phenology](#)

REFLECTED SOLAR RADIATION SENSORS, MULTIANGLE IMAGING

David J. Diner

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

Multangular imaging; Multidirectional imaging; Multiple view angle imaging

Definitions

Sequential multiangle imaging. The acquisition of images of a scene at different view zenith angles (the angle between the local vertical and the direction to the observer) as a result of observing the scene on different occasions, for example, on different orbits of a satellite. In this case, the time separation between views may be hours, days, or weeks.

Simultaneous multiangle imaging. The acquisition of images of a scene at different view zenith angles during a single overpass, obtained by reorienting the sensor in the along-track direction (the direction parallel to the sensor's flight direction) or by using multiple sensors pointed at different along-track angles to observe the target repeatedly as the view zenith angle changes. Here, the time separation between views is several minutes.

Introduction

Clouds, aerosol layers, vegetation canopies, soils, snow or ice fields, and other natural scenes reflect solar radiation into angular reflectance patterns that depend on the angle at which the scene is illuminated or viewed. The Bidirectional Reflectance Distribution Function (BRDF) (Nicodemus et al., 1977; Schaepman-Strub et al., 2006) is the ratio of the radiance reflected by a target as a function of view direction, divided by the irradiance illuminating the target at a single incidence angle. The magnitude and angular shape of the BRDF is largely governed by the composition, density, and geometric structure of the reflecting medium.

Many satellite instruments having a wide image swath, for example, the multispectral MODerate-resolution Imaging Spectroradiometer (MODIS) on the National Aeronautics and Space Administration (NASA) Terra and Aqua satellites, acquire multiangle data according to the sequential approach defined above. For many applications, the simultaneous multiangle measurement strategy is required, and a number of satellite sensors have been

designed with this objective in mind. Examples of such instruments are described below and provide the focus of this discussion. As noted by Diner et al. (2005), the interpretation of multiangle information can be categorized into two main strategies: geometric and radiometric. The first involves real or apparent differences in the location of observed objects with changing angle of view, resulting from stereoscopic parallax (displacement dependent upon distance from the observer) or actual motion of the target during the time interval between views. The second refers to changes in the brightness, color, contrast, or polarization of the reflected light as a function of view angle.

Satellite sensors

Several Earth-orbiting instruments have been designed to make simultaneous multiangle observations at visible and shortwave infrared wavelengths (see Table 1). Airborne counterparts are associated with some of these instruments. In addition, airborne sensors such as the Cloud Absorption Radiometer (CAR) with 13 bands between 503 and 2,289 nm (King et al., 1986) (later upgraded to 14 bands between 340 and 2,302 nm; Gatebe et al., 2003) and the Advanced Solid-State Array Spectroradiometer (ASAS) with 29 bands from 465 to 871 nm (Irons et al., 1991) have provided a legacy on the information content of multiangle imagery.

- The United Kingdom's Along-Track Scanning Radiometer (ATSR)-1 (Delderfield et al., 1986; Mutlow et al., 1994) was launched on the European Space Agency (ESA) Earth remote sensing satellite (ERS)-1 in 1991. It was primarily designed to retrieve global sea-surface temperatures. The only solar reflective channel in this instrument was a band at 1.6 μm . An enhanced version, ATSR-2, was equipped with additional visible and near-infrared channels and was launched in 1995 aboard ERS-2. In 2002, the Advanced ATSR (AATSR) was launched on ENVISAT. Each of these instruments contains a conical scanning radiometer that provides curved swaths at two measurement angles, one near nadir and the other at an oblique forward-viewing angle (<http://www.atsr.rl.ac.uk/>).
- Contact with the ENVISAT spacecraft was lost in spring 2012, and ESA declared an end to the mission.
- The Polarization and Directionality of the Earth's Reflectances (POLDER) instrument (Deschamps et al., 1994), developed by the French Centre National d'Etudes Spatiales (CNES), was launched aboard the Japanese Advanced Earth Observing Satellite (ADEOS) in 1996 and acquired 8 months of global data. A follow-on instrument, POLDER-2, was launched on ADEOS-2 in 2002 and acquired data for nearly a year. Both missions were cut short by spacecraft failures. A third POLDER instrument was launched in 2004 on the French PARASOL satellite as part of the polar-orbiting "A-train," a constellation of satellites in similar orbits to acquire near-simultaneous observations. In late 2011, PARASOL separated from the A-train orbit, but the instrument is still collecting data. POLDER uses a wide-angle imaging system and an area array detector to acquire measurements at a multitude of along-track and cross-track angles (<http://www.icare.univ-lille1.fr/parasol/>). An airborne version of this instrument, AirPOLDER, also exists (<http://loag5-ct13.univ-lille1.fr/AirPOLDER/data.html>).
- The Multiangle Imaging SpectroRadiometer (MISR) (Diner et al., 1998a), developed in the United States, was launched aboard the NASA Terra spacecraft, part of the Earth Observing System, in late 1999. The MISR instrument uses nine separate cameras to acquire data at nine discrete along-track observation angles. Images are acquired globally in push broom fashion, with four line arrays in each camera filtered to four visible and near-infrared spectral bands (<http://mISR.jpl.nasa.gov>). An airborne instrument making use of a single, gimbaled camera to acquire multiangle imagery – AirMISR – has also been used for aerosol, cloud, and surface investigations (Diner et al., 1998b; <http://mISR.jpl.nasa.gov/Mission/airMISR.html>).
- The US Department of Energy's Multispectral Thermal Imager (MTI) (Szymanski and Weber, 2005) contains 15 spectral bands of which 10 are at solar reflective wavelengths. MTI was launched in 2000. The satellite has an agile-pointing capability and multiangle imagery

Reflected Solar Radiation Sensors, Multiangle Imaging, Table 1 Earth-orbiting multiangle remote sensing imagers

	(A)ATSR	POLDER	MISR	MTI	CHRIS	CIPS
Number of along-track view angles	2	14	9	2	5	7
Maximum view angle (at Earth's surface)	56° (forward only)	60° (forward + backward)	70° (forward + backward)	60° (typically backward)	55° (forward + backward)	71° (forward + backward)
Shortwave spectral bands	555, 659, 865, 1610 nm	443 ^a , 490, 565, 670 ^a , 763, 765, 865 ^a , 910 nm (^a polarized)	446, 558, 672, 866 nm	484, 558, 650, 810, 874, 940, 1,015, 1,376, 1,646, 2,224 nm	Nominally 18 channels from 415 to 1,050 nm	265 nm
Footprint (nadir)	1 × 1 km ²	6 × 7 km ²	275 × 275 m ² , 1.1 × 1.1 km ²	5 × 5 m ² , 20 × 20 m ²	17 × 17 m ²	1 × 2 km ²
Swath width	512 km	2,200 km	400 km	12 km	13 km	950 km

is acquired by reorienting the satellite during a target overpass. Due to this strategy and the sensor's narrow field of view, MTI is a targeting instrument (<http://www.lanl.gov/orgs/isr/isr2/>).

- The Compact High Resolution Imaging Spectrometer (CHRIS) instrument (Barnsley et al., 2004) was developed in the United Kingdom and launched in 2001 on an agile spacecraft known as PROBA (Project for On-Board Autonomy), under the auspices of ESA. Like MTI, CHRIS is a targeting instrument and acquires multiangle imagery by pointing the PROBA spacecraft (<http://earth.esa.int/missions/thirdpartymission/proba.html>).
- The Cloud Imaging and Particle Size Experiment (CIPS) (McClintock et al., 2009), developed in the United States, was launched aboard NASA's Aeronomy of Ice in the Mesosphere (AIM) mission in 2007. CIPS uses an array of four cameras in a "+" configuration, two pointing forward and backward, and two pointing to either side, each with a $44^\circ \times 44^\circ$ field of view. Observations are acquired in a single ultraviolet spectral band to measure the physical characteristics of polar mesospheric clouds (PMCs, also known as noctilucent clouds to ground-based observers) (<http://aim.hamptonu.edu/instrmt/cips.html>).

Example applications

Aerosols (liquid or solid airborne particulates having sizes ranging from less than $0.1 \mu\text{m}$ to more than $10 \mu\text{m}$) and clouds (consisting of larger-sized liquid water or ice particles) play critical roles in our climate system. Whether a particular cloud or aerosol cools or warms the surface depends on the physical properties of the particles, as well as their heights. For aerosols, injection height into the atmosphere also determines how far the particles can be transported away from their source, with ramifications for downwind air quality.

The angular distribution of scattering from particles is different for spherical (liquid) cloud or aerosol droplets and among various irregularly shaped (e.g., dust) or crystal (ice) forms. As a result, multiangle imagery has proven valuable in distinguishing cloud and aerosol types characterized by the particle shape, one of several factors affecting climate impact. Data from ATSR-2, POLDER, and MISR have been used to explore ice crystal habit in cirrus clouds (Baran et al., 1999; Chepfer et al., 2001; McFarlane and Marchand, 2008). Scattering phase functions from CIPS were used to derive particle sizes of polar mesospheric clouds (Bailey et al., 2009). Additional information on cloud particle size as well as shape is contained in multiangular polarization data (Parol et al., 2004). Kalashnikova and Kahn (2006) demonstrated the ability to distinguish among granular, platelike, and spheroidal dust particles using MISR data over ocean. Discriminating between nonspherical and spherical particles is also useful in separating dust from fine particle pollution over land. Liu et al. (2007) showed improved correlations between column aerosol amount and the concentration

of near-surface airborne particulate matter with diameters less than $2.5 \mu\text{m}$ ($\text{PM}_{2.5}$, a regulated air pollutant) by using multiangle data.

Over land, accurate satellite aerosol retrievals are complicated by the large spatial variability in surface reflectance. Bright deserts and urban areas are major aerosol source regions, and separating the surface and atmospheric contributions to the observed top-of-atmosphere radiances is challenging. ATSR-2/AATSR and MISR use multiangle observations to solve this problem, making use of the enhanced atmospheric signal at oblique view angles and differences in the manner in which atmospheric and surface layers scatter radiation (Flowerdew and Haigh, 1996; Martonchik et al., 2002). Enhanced retrieval accuracy is obtained at certain view geometries, a subject explored using MTI (Chylek et al., 2003). Differences in the angular reflectance signature of clouds and snow/ice have proven valuable in improving cloud detection capability in polar regions (Di Girolamo and Wilson, 2003; Shi et al., 2007).

Many satellite stereoscopic imagers designed to retrieve land surface topography have been placed into Earth's orbit. Either the sequential or simultaneous viewing strategy can be used for this purpose. For remote sensing of highly dynamic targets such as clouds, however, the simultaneous multiangle strategy is required. Cloud stereo imagery can be acquired with more than one geostationary satellite viewing the targets at different angles (Seiz et al., 2007), but these equatorial-orbiting satellites do not observe polar latitudes. Alternatively, stereo observations can be obtained from a single instrument in polar orbit by acquiring data at more than one along-track view angle. Stereoscopic imaging from MISR and ATSR has been used to capture global cloud-top heights (CTH) (Moroney et al., 2002; Prata and Turner, 1997; Naud et al., 2006). Unlike thermal infrared-based techniques, which are based on measurements of cloud-top temperature, the stereoscopic approach is purely geometric, thereby making it capable of providing cloud height climatologies independent of confounding factors such as temperature inversions (Garay et al., 2008; Harshvardhan et al., 2009). Efficient, automated computational methods have been developed to "match" the images of clouds seen from different view angles (Muller et al., 2002); knowledge of the spacecraft position and view angles is then used to obtain CTH. Stereo imaging using at least three angles has made possible a novel technique for simultaneously retrieving height-resolved cloud-tracked velocity (Horváth and Davies, 2001; Zong et al., 2002). Imaging at multiple view angles helps quantify the effect of 3D cloud morphology on reflected radiation (Horváth and Davies, 2004; Cornet and Davies, 2008). Stereo imaging is also able to determine aerosol plume heights. Statistical studies of wildfire plumes in North America between 2002 and 2007 using MISR show that (a) a significant fraction of fires inject smoke into the free troposphere, (b) smoke tends to concentrate in stable layers within the atmosphere, and (c) there is a pronounced seasonal cycle of plume

heights over certain vegetation types (Kahn et al., 2008; Val Martin et al., 2010).

Angular reflectances of surface covers contain information about their physical structure (e.g., Widlowski et al., 2004). Ice surface roughness provides an indicator of snow accumulation, surface melt, and wind ablation, and studies using the sensitivity of multiangle observations to surface roughness show great potential for distinguishing glacial zones and identifying different types of sea ice, which are indicators of climate change (Nolin et al., 2002; Nolin and Payne, 2007). For terrestrial vegetation, the canopy and understory density and morphology (e.g., height and width) of the plants and trees impact photosynthetic efficiency, nutrient cycling, suitability as wildlife habitats, fresh water availability, and fire risk. Multiangle data are particularly useful for separating canopy and understory/ground reflectance (Canisius and Chen, 2007; Pinty et al., 2008), which is particularly significant for bright surfaces (i.e., snow or desert). CHRIS/PROBA and MISR have been used for remote sensing of desert shrub cover (Chopping et al., 2003, 2008a), which has been expanding during the last century in the southwestern USA at the expense of grasslands. Geometric-optical methods for using multiangle data to retrieve plant number density, mean radius, crown height, and crown shape in forests are also being developed (e.g., Chopping et al., 2008b). One technique that holds significant promise is the use of active lidar, which provides more direct measurements of tree heights, in conjunction with passive multiangular imagery to retrieve canopy structure (Kimes et al., 2006; Schull et al., 2007).

Conclusions

Multiangle imaging is a rich source of information regarding the micro- and macrophysical structure of different types of aerosols, cloud forms, and surface covers. Combined with stereoscopic techniques, it enables construction of 3D scene models and estimation of the total amount of sunlight reflected by Earth's diverse environments. As with any imaging technology, the range of applications is myriad. Just a few representative examples have been presented here. Many other uses exist, and additional ones likely remain to be discovered.

Acknowledgment

The research to prepare this contribution was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Bibliography

- Bailey, S. M., Thomas, G. E., Rusch, D. W., Merkel, A. W., Jeppesen, C. D., Carstens, J. N., Randall, C. E., McClintock, W. E., and Russell, J. M., III, 2009. Phase functions of polar mesospheric cloud ice as observed by the CIPS instrument on the AIM satellite. *Journal of Atmospheric and Terrestrial Physics*, **71**, 373–380.
- Baran, A. J., Watts, P. D., and Francis, P. N., 1999. Testing the coherence of cirrus microphysical and bulk properties retrieved from dual-viewing multispectral satellite radiance measurements. *Journal of Geophysical Research*, **104**, 31673–31683.
- Barnsley, M. J., Settle, J. J., Cutter, M., Lobb, D., and Teston, F., 2004. The PROBA/CHRIS mission: a low-cost smallsat for hyperspectral, multi-angle, observations of the Earth surface and atmosphere. *IEEE Transactions on Geoscience & Remote Sensing*, **42**, 1512–1520.
- Canisius, F., and Chen, J. M., 2007. Retrieving vegetation background reflectance from multi-angle imaging spectroradiometer (MISR) data. *Remote Sensing of Environment*, **107**, 312–321.
- Chepfer, H., Goloub, P., Riedi, J., Haan, J. F. D., Hovenier, J. W., and Flamant, P. H., 2001. Ice crystal shapes in cirrus clouds derived from POLDER/ADEOS-1. *Journal of Geophysical Research*, **106**, 7955–7966.
- Chopping, M., Su, L., Laliberte, A., Rango, A., Peters, D. P. C., and Kollikkathara, N., 2003. Mapping shrub abundance in desert grasslands using geometric-modeling and multi-angle remote sensing with CHRIS/Proba. *Remote Sensing of Environment*, **104**, 62–73.
- Chopping, M., Su, L., Rango, A., Martonchik, J. V., Peters, D. P. C., and Laliberte, A., 2008a. Remote sensing of woody shrub cover in desert grasslands using MISR with a geometric-optical canopy reflectance model. *Remote Sensing of Environment*, **112**, 19–34.
- Chopping, M., Moisen, G., Su, L., Laliberte, A., Rango, A., Martonchik, J. V., and Peters, D. P. C., 2008b. Large area mapping of southwestern forest crown cover, canopy height, and biomass using MISR. *Remote Sensing of Environment*, **112**, 2051–2063.
- Chylek, P., Henderson, B., and Mishchenko, M., 2003. Satellite based retrieval of aerosol optical thickness: the effect of sun and satellite geometry. *Geophysical Research Letters*, **30**, doi:10.1029/2003GL016917.
- Cornet, C., and R. Davies, 2008. Use of MISR measurements to study the radiative transfer of an isolated convective cloud: implications for cloud optical thickness retrieval. *Journal of Geophysical Research*, **113**, Art. No. D04202.
- Delderfield, J., Llewellyn-Jones, D. T., Bernard, R., de Javel, Y., Williamson, E. J., Mason, I., Pick, D. R., and Barton, I. J., 1986. The along track scanning radiometer (ATSR) for ERS-1. *Proceedings of SPIE*, **589**, 114–120.
- Deschamps, P.-Y., Bréon, F. M., Leroy, M., Podaire, A., Bricaud, A., Buriez, J.-C., and Sèze, G., 1994. The POLDER mission: instrument characteristics and scientific objectives. *IEEE Transactions on Geoscience & Remote Sensing*, **32**, 598–615.
- Di Girolamo, L., and Wilson, M. J., 2003. A first look at band-differenced angular signatures for cloud detection from MISR. *IEEE Transactions on Geoscience & Remote Sensing*, **41**, 1730–1734.
- Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Muller, J.-P., Myneni, R., Sellers, P. J., Pinty, B., and Verstraete, M. M., 1998a. Multi-angle imaging spectroradiometer (MISR) description and experiment overview. *IEEE Transactions on Geoscience & Remote Sensing*, **36**, 1072–1087.
- Diner, D. J., Barge, L. M., Bruegge, C. J., Chrien, T. G., Conel, J. E., Eastwood, M. L., Garcia, J. D., Hernandez, M. A., Kurzweil, C. G., Ledebauer, W. C., Pignatano, N. D., Sarture, C. M., and Smith, B. G., 1998b. The airborne multi-angle imaging spectroradiometer (AirMISR): instrument description and first results. *IEEE Transactions on Geoscience & Remote Sensing*, **36**, 1339–1349.
- Diner, D. J., Braswell, B. H., Davies, R., Gobron, N., Hu, J., Jin, Y., Kahn, R. A., Knyazikhin, Y., Loeb, N., Muller, J.-P., Nolin, A. W., Pinty, B., Schaaf, C. B., Seiz, G., and Stroeve, J., 2005. The value of multiangle measurements for retrieving structurally

- and radiatively consistent properties of clouds, aerosols, and surfaces. *Remote Sensing of Environment*, **97**, 495–518.
- Flowerdew, R. J., and Haigh, J. D., 1996. Retrieval of aerosol optical thickness over land using the ATSR-2 dual-look satellite radiometer. *Geophysical Research Letters*, **23**, 351–354.
- Garay, M. J., de Szoek, S. P., and Moroney, C. M., 2008. Comparison of marine stratocumulus cloud top heights in the Southeastern Pacific retrieved from satellites with coincident ship-based observations. *Journal of Geophysical Research*, **113**, D18204.
- Gatebe, C. K., King, M. D., Platnick, S., Arnold, G. T., Vermote, E. F., and Schmid, B., 2003. Airborne spectral measurements of surface-atmosphere anisotropy for several surfaces and ecosystems over southern Africa. *Journal of Geophysical Research*, **108**, doi:10.1029/2002JD002397.
- Harshvardhan, Zhao, G., Di Girolamo, L., and Green, R. N., 2009. Satellite-observed location of stratocumulus cloud-top heights in the presence of strong inversions. *IEEE Transactions on Geoscience and Remote Sensing*, **47**, 1421–1428.
- Horváth, Á., and Davies, R., 2001. Simultaneous retrieval of cloud motion and height from polar-orbiter multiangle measurements. *Geophysical Research Letters*, **28**, 2915–2918.
- Horváth, Á., and Davies, R., 2004. Anisotropy of water cloud reflectance: a comparison of measurements and 1D theory. *Geophysical Research Letters*, **31**, L01102, doi:10.1029/2003GL018386.
- Irons, J. R., Ranson, K. J., Williams, D. L., Irish, R. R., and Huegel, F. G., 1991. An off-nadir-pointing imaging spectroradiometer for terrestrial ecosystem studies. *IEEE Transactions on Geoscience & Remote Sensing*, **29**, 66–74.
- Kahn, R., Chen, Y., Nelson, D. L., Leung, F.-Y., Li, Q., Diner, D. J., and Logan, J. A., 2008. Wildfire smoke injection heights: two perspectives from space. *Geophysical Research Letters*, **35**, L04809, doi:10.1029/2007GL032165.
- Kalashnikova, O. V., and Kahn, R., 2006. Ability of multiangle remote sensing observations to identify and distinguish mineral dust types: part 2. Sensitivity over dark water. *Journal of Geophysical Research*, **111**, D11207, doi:10.1029/2005JD006756.
- Kimes, D. S., Ranson, K. J., Sun, G., and Blair, J. B., 2006. Predicting lidar measured forest vertical structure from multi-angle spectral data. *Remote Sensing of Environment*, **100**, 503–511.
- King, M. D., Strange, M. G., Leone, P., and Blaine, L. R., 1986. Multiwavelength scanning radiometer for airborne measurements of scattered radiation within clouds. *Journal of Atmospheric and Oceanic Technology*, **3**, 513–522.
- Liu, Y., Koutrakis, P., Kahn, R., Turquety, S., and Yantosca, R. M., 2007. Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: part 2 – a case study. *Journal of the Air & Waste Management Association*, **57**, 1360–1369.
- Martonchik, J. V., Diner, D. J., Crean, K. A., and Bull, M. A., 2002. Regional aerosol retrieval results from MISR. *IEEE Transactions on Geoscience & Remote Sensing*, **40**, 1520–1531.
- McClintock, W. E., Rusch, D. W., Thomas, G. E., Merkel, A. W., Lankton, M. R., Drake, V. A., Bailey, S. M., and Russell, J. M., III, 2009. The cloud imaging and particle size experiment on the aeronomy of ice in the mesosphere mission: instrument concept, design, calibration, and on-orbit performance. *Journal of Atmospheric and Terrestrial Physics*, **71**, 340–355.
- McFarlane, S. A., and Marchand, R. T., 2008. Analysis of ice crystal habits derived from MISR and MODIS observations over the ARM Southern Great Plains site. *Journal of Geophysical Research*, **113**, D07209, doi:10.1029/2007JD009191.
- Moroney, C., Davies, R., and Muller, J.-P., 2002. Operational retrieval of cloud-top heights using MISR data. *IEEE Transactions on Geoscience & Remote Sensing*, **40**, 1541–1546.
- Muller, J.-P., Mandanayake, A., Moroney, C., Davies, R., Diner, D. J., and Paradise, S., 2002. MISR stereoscopic image matchers: techniques and results. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 1547–1559.
- Mutlow, C. T., Zavody, A. M., Barton, I. J., and Llewellyn-Jones, D. T., 1994. Sea-surface temperature-measurements by the along-track scanning radiometer on the ERS-1 satellite – early results. *Journal of Geophysical Research*, **99**, 22575–22588.
- Naud, C., Muller, J.-P., and Clothiaux, E. E., 2006. Assessment of multispectral ATSR2 stereo cloud-top height retrievals. *Remote Sensing of Environment*, **104**, 337–345.
- Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., and Limperis, T., 1977. *Geometrical Considerations and Nomenclature for Reflectance*. Washington, DC: National Bureau of Standards, U.S. Department of Commerce. NBS Monograph, Vol. 160.
- Nolin, A. W., and Payne, M., 2007. Classification of glacier zones in western Greenland using albedo and surface roughness from the multi-angle imaging spectroradiometer (MISR). *Remote Sensing of Environment*, **107**, 264–275.
- Nolin, A. W., Fetterer, F. M., and Scambos, T. A., 2002. Surface roughness characterizations of sea ice and ice sheets: case studies with MISR data. *IEEE Transactions on Geoscience & Remote Sensing*, **40**, 1605–1615.
- Parol, F., Buriez, J. C., Vanbouce, C., Riedi, J., Labonnote, L. C., Doutriaux-Boucher, M., Vesperini, M., Sèze, G., Couvert, P., Viollier, M., and Bréon, F.-M., 2004. Capabilities of multi-angle polarization cloud measurements from satellite: POLDER results. *Advances in Space Research*, **33**, 1080–1088.
- Pinty, B., Laverne, T., Kaminski, T., Aussedat, O., Giering, R., Gobron, N., Taberner, M., Verstraete, M. M., Vobbeck, M., and Widlowski, J.-L., 2008. Partitioning the solar radiant fluxes in forest canopies in the presence of snow. *Journal of Geophysical Research*, **113**, D04104, doi:10.1029/2007JD009096.
- Prata, A. J., and Turner, P. J., 1997. Cloud top height determination using ATSR data. *Remote Sensing of Environment*, **59**, 1–13.
- Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S., and Martonchik, J. V., 2006. Reflectance quantities in optical remote sensing – definitions and case studies. *Remote Sensing of Environment*, **103**, 27–42.
- Schull, M. A., Ganguly, S., Samanta, A., Huang, D., Shabanov, N. V., Jenkins, J. P., Chiu, J. C., Marshak, A., Blair, J. B., Myneni, R. B., and Knyazikhin, Y., 2007. Physical interpretation of the correlation between multi-angle spectral data and canopy height. *Geophysical Research Letters*, **34**, L18405, doi:10.1029/2007GL031143.
- Seiz, G., Tjemkes, S., and Watts, P., 2007. Multiview cloud-top height and wind retrieval with photogrammetric methods: application to Meteosat-8 HRV observations. *Journal of Applied Meteorology and Climatology*, **46**, 1182–1195.
- Shi, T., Clothiaux, E. E., Yu, B., Braverman, A. J., and Groff, D. N., 2007. Detection of daytime arctic clouds using MISR and MODIS data. *Remote Sensing of Environment*, **107**, 172–184.
- Szymanski, J., and Weber, P., 2005. Multispectral thermal imager: mission and applications overview. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 1943–1949.
- Val Martin, M., Logan, J. A., Kahn, R. A., Leung, F.-Y., Nelson, D. L., and Diner, D. J., 2010. Smoke injection heights from fires in North America: analysis of 5 years of satellite observations. *Atmospheric Chemistry and Physics*, **10**, 1491–1510.
- Widlowski, J.-L., Pinty, B., Gobron, N., Verstraete, M. M., Diner, D. J., and Davis, A. B., 2004. Canopy structure parameters derived from multi-angular remote sensing data for terrestrial carbon studies. *Climatic Change*, **67**, 403–415.
- Zong, J., Davies, R., Muller, J.-P., and Diner, D. J., 2002. Photogrammetric retrieval of cloud advection and top height from the Multi-angle imaging spectroradiometer (MISR). *Photogrammetric Engineering and Remote Sensing*, **68**, 821–829.

Cross-references

Aerosols
 Cloud Properties
 Optical/Infrared, Scattering by Aerosols and Hydrometeors
 Reflected Solar Radiation Sensors, Polarimetric

REFLECTED SOLAR RADIATION SENSORS, POLARIMETRIC

David J. Diner
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Synonyms

Stokes measurements

Definitions

Stokes vector. A set of four parameters describing the polarization state of a beam of light, named for the Irish mathematical physicist George Gabriel Stokes (1819–1903). As an electromagnetic wave propagates, the orientation of the tip of the electric field vector traces out an ellipse. Specific manifestations of the polarization ellipse are linear polarization, where the electric field vibrates in a single plane, and circular polarization. The Stokes vector consists of intensity, I ; the excess of horizontally over vertically polarized light, Q ; the excess of light polarized at 45° over 135° , U ; and the excess of right-handed over left-handed circular polarization, V . “Handedness” describes the direction in which the electric field vector rotates.

Degree of polarization. The ratio of polarized to total intensity is equal to $\sqrt{Q^2 + U^2 + V^2}/I$. The degree of linear polarization (DOLP) is equal to $\sqrt{Q^2 + U^2}/I$, and the degree of circular polarization (DOCP) is equal to V/I .

Angle of polarization. The orientation of the major axis of the polarization ellipse. In terms of the Stokes components, the angle of polarization equals to $0.5 \arctan(U/Q)$.

Polarimetry. Measurement of the Stokes vector, typically involving the use of polarization *analyzers*, devices that preferentially transmit certain states of polarization, and *retarders* or *phase plates*, which alter the shape of the polarization ellipse. *Division-of-amplitude* polarimeters optically divide the incoming light and direct the resulting beams toward multiple analyzers and detectors. *Division-of-aperture* polarimeters employ multiple analyzers operating side-by-side. *Snapshot* devices encode the polarization state within a spatially varying signal recorded on an area array detector. *Spectrum* channeling approaches use thick birefringent crystals plus a polarizer and spectrometer to encode the Stokes components in fringes superimposed on the recorded spectrum. A *time-multiplexed* polarimeter uses a rotating analyzer or

retarder, or an electro-optical retardance modulator, and the measurements are acquired sequentially.

Introduction

Sunlight incident on the Earth system is unpolarized, but scattering by clouds, aerosols, and surfaces in general polarizes the light. Polarimeters of myriad designs, operating in various wavelength intervals, have been used in a host of scientific applications. A classic example of the application of polarimetry to planetary astronomy was the identification of sulfuric acid clouds in the upper atmosphere of Venus (Hansen and Hovenier, 1974). This entry presents examples of Earth remote sensing polarimeters that record reflected sunlight photographically or photometrically from suborbital and orbital platforms. With the exception of a brief discussion on rocket-borne observations of mesospheric clouds, the scope is confined to sensors viewing the Earth at angles between the downward (nadir) and horizontal (limb) directions.

Polarimetric remote sensing from balloons

Photopolarimetry of earthlight was obtained from high-altitude balloons over Arizona and New Mexico by researchers at the University of California, Los Angeles (Rao and Sekera, 1967; Rao, 1969). Photomultiplier tube-based instruments with a 3° field of view (FOV) were scanned on either side of nadir. Initial measurements, from 13 km altitude, used a rotating half-wave retarder, a fixed polarizing prism, and spectral filters at 327, 398, 500, and 612 nm. Later observations (from 28 km) used 362, 401, 501, and 599 nm filters and a rotating calcite prism for polarimetric modulation. Balloon-borne (20 km altitude) polarimetry over southwestern France was obtained by Herman et al. (1986) at 850 and 1,650 nm using a rotating Polaroid analyzer and a germanium photodiode. Horizontal scans obtained when the Sun was close to the horizon provided evidence of spherical submicron particles consistent with established stratospheric aerosol models. Downward-looking observations of tropospheric aerosols, clouds, and the ocean surface were later obtained by adding a scan mirror (Deuzé et al., 1989).

The three commonly known neutral points in the polarization of skylight – Arago, Babinet, and Brewster, named after their discoverers – are locations where the degree of polarization vanishes, typically displaced tens of degrees from the solar or antisolar directions. Horváth et al. (2002) confirmed the existence of a suspected fourth neutral polarization point below the antisolar direction (thus invisible from the ground). This was accomplished by flying a Nikon F801 film camera equipped with a fisheye lens and linear polarizers oriented at 0° , 45° , and 90° (Gál et al., 2001) in a low-altitude (0.8–3.5 km) hot air balloon. The data were collected during two flights over central Hungary.

Polarimetric remote sensing from rockets

Noctilucent or polar mesospheric clouds consist of tiny ice particles that reside in the Earth's upper atmosphere at altitudes near 80 km. Remote sensing from suborbital rockets enables determination of their light-scattering properties and vertical distribution. Witt et al. (1976) at the University of Stockholm used a photomultiplier-based instrument operating at 256 and 536 nm to derive a particle size upper bound of ~ 50 nm. A two-color (410 and 540 nm) polarimeter from the Dudley Observatory (New York) implied somewhat larger sizes (Tozer and Beeson, 1974). More recently, the Ultraviolet Imaging Polarimeter (UVIP) built by the University of Colorado (Lawrence et al., 1994) provided rocket-borne limb imaging using a Reticon detector array. Data at polarizer orientations of 0° , 45° , and 90° were acquired at 265 nm.

Polarimetric remote sensing from airplanes

Among the earliest non-imaging aircraft polarimeters was a Grumman instrument deployed on a Piper Tri-Pacer (Egan, 1968). The instrument had a 1° FOV and was equipped with filters at 533 and 1,000 nm and employed a rotating Polaroid analyzer. Initial results over forest discriminated red pine and hardwood trees. Hariharan (1969) described an airborne polarimeter for atmospheric radiation studies consisting of polarizing (Glan-Thompson) prisms, four blue-green spectral filters, and a photomultiplier tube. It flew in the National Aeronautics and Space Administration (NASA) CV-990 aircraft and made observations over a uniform cloud deck. A University of Arizona instrument (Fernald et al., 1969) with a Polaroid analyzer oriented at 0° , 90° , and 120° observed sparsely vegetated desert terrain at 429 nm from 300 to 600 m altitude. DOLP varied from 3 % to 18 %, depending on scattering angle, and the angle of polarization was generally perpendicular to the scattering plane. The University of Arizona's AEROPOL instrument (Coffeen et al., 1975) flew on the NASA CV-990. Observations within a 1.5° FOV were acquired between 1,100 and 3,500 nm using a rotating wire-grid analyzer to study cloud and aerosol polarization. Comparisons with theoretical computations showed sensitivity to cloud phase and particle size distribution (Hansen and Coffeen, 1974).

Early airborne imaging polarimeters included the Grumman Digital Photometric Mapper (DPM) (Halajian and Hallock, 1972) and an instrument developed at the University of Cologne (Prosch et al., 1983). The DPM used an image dissector tube and required multiple passes to obtain different polarization orientations. An inverse relationship between water turbidity and DOLP was demonstrated. The Cologne instrument contained three Plumbicon-based video cameras mounted side-by-side with polarizing filters at 0° , 60° , and 120° , spectrally filtered to a single band at 550 nm. Data collected over the Olefalsperre reservoir indicated a sensitivity to surface wind speed.

The Airborne Polarization and Directionality of Earth's Reflectances (AirPOLDER) instrument, built by the Laboratoire d'Optique Atmosphérique at the Université de Lille became operational in June 1990 (Deuzé et al., 1993). A rotating wheel provides spectral and polarimetric filter selection in discrete bands between 443 and 910 nm. A silicon charge-coupled device (CCD) detector array and a wide-angle lens provide imagery over a $\pm 41^\circ \times \pm 51^\circ$ FOV. Successive images obtained during flight enable each point to be observed at 12 view angles. More recently, the Lille group developed a single-view-angle aerosol polarimeter, MICROPOL, which operates in the visible/near-infrared (VNIR) and shortwave-infrared (SWIR) (Waquet et al., 2005). Three separate optical systems at each wavelength contain polarizers in orientations separated by 60° . A new instrument based on the AirPOLDER design, the Observing System Including polaRization in the solar Infrared Spectrum (OSIRIS), uses separate VNIR and SWIR optical systems to acquire data at 440, 490, 670, 763, 765, 865, 910, 940, 1,020, 1,240, 1,365, 1,600, and 2,200 nm (Auriol et al., 2008). It has been tested aboard a French research aircraft flying at 10 km altitude.

Several of the instruments described in this entry are *incomplete polarimeters* (Chipman, 1994) that measure linear polarization in just two perpendicular orientations, typically providing Q but not U . One such sensor was developed at NASA Ames Research Center to study oil slicks (Millard and Arvesen, 1972). Measurements at 380 nm over controlled spills were acquired with polarizers parallel and perpendicular to the flight direction of a Cessna 401. The difference between the two measurements increased by 25 % over the oil-contaminated water. Another Ames sensor was developed for the high-altitude ER-2 aircraft (Hildum and Spinhirne, 1992). It used a commercial silicon CCD camera to acquire data in support of cloud observation campaigns. Colorado State University's Scanning Spectral Polarimeter (SSP) also included channels for left and right circular polarization (Stephens et al., 2000).

The Research Scanning Polarimeter (RSP), developed by SpecTIR Corporation and operated by the NASA Goddard Institute for Space Studies (Cairns et al., 2003), was designed for high polarimetric accuracy (<0.2 % uncertainty in DOLP). RSP observes in nine spectral bands (410, 470, 555, 670, 865, 960, 1,590, 1,880, and 2,250 nm). Refractive telescopes measure the Stokes parameters I , Q , and U simultaneously. A rotating scan mirror assembly sweeps the telescope FOVs to enable data collection at $\pm 60^\circ$ from nadir. The two mirrors in the scan assembly are oriented such that any polarization introduced by the first mirror is compensated by the second. RSP has flown on different aircraft in a variety of field campaigns and has been used in a cross-track scanning mode to provide imagery as well as in an along-track scanning mode to obtain polarimetric angular signatures of aerosols, clouds, and land surfaces.

The Airborne Multi-Spectral Sunphoto- and Polarimeter (AMSSP) developed at the Freie Universität Berlin is

designed to operate in conjunction with an optical scan head that enables multidirectional observations of both the upwelling and downwelling hemispheres (Ruhtz et al., 2009). AMSSP contains a combination of analyzers and retarders to measure all four components of the Stokes vector in the visible and ultraviolet. An engineering model of the instrument has flown aboard a British research aircraft in field campaigns aimed at cloud and aerosol studies.

The Airborne Optical Spotlight System-Multispectral Polarimeter (AROSS-MSP), built by Areté Associates in Virginia, is a 12-camera system containing one camera for each of four spectral bands (blue, green, red, near-infrared) in each of three polarization orientations (Hooper et al., 2009). It is an upgrade of an earlier 9-camera system that included the three visible bands only. Designed to measure suspended material transport in rivers and near-shore ocean environments, AROSS-MSP has flown on a Twin Otter aircraft and was able to track water current velocities using high spatial resolution time-lapse imaging.

Several aircraft polarimeters use nonconventional approaches. Utah State University's Hyperspectral Imaging Polarimeter (HIP) employed ferroelectric liquid crystal (FLC) retarders to rapidly switch between different polarization orientations (Jensen and Peterson, 1998). HIP was designed for high spatial and spectral resolution polarimetry at 2,700 nm to determine cloud particle phase. The Hyperspectral Polarimeter for Aerosol Retrievals (HySPAR) developed by Aerodyne Research (Jones et al., 2006) is a spectrum channeling imaging spectrometer operating at 480–960 nm. It has participated in field campaigns aboard the NASA King Air B200 aircraft. The Airborne Multiangle SpectroPolarimetric Imager (AirMSPI), built by the Jet Propulsion Laboratory in collaboration with the University of Arizona, uses photoelastic modulators (PEMs) to provide rapid time multiplexing of the Stokes components (Diner et al., 2007, 2010). AirMSPI obtains intensity images between 355 and 935 nm and polarization images at 470, 660, and 865 nm. It is mounted on a gimbal to acquire imagery at multiple along-track view angles and has been flying aboard NASA's ER-2 high-altitude aircraft since 2010 (Diner et al., 2011). Another imaging design for atmospheric remote sensing makes use of a Philips-type prism (Fernandez-Borda et al., 2009). Incoming light is split into three optical channels for simultaneous measurement in three polarizer orientations. The Visible-regime Polarimetric Imager developed at TRW (Barter et al., 2003) uses a prism to split the beam into four optical channels. CCD arrays are used in each channel and a combination of polarizers and a quarter-wave retarder enables imaging all four Stokes components in a single mid-visible spectral band. Airborne images of DOLP and DOCP were acquired over agricultural fields.

The rainbow and cloudbow are scattering phenomena resulting from refraction and reflection of light within spherical droplets. Light in the bow is polarized and its

variation with angle depends on the droplet size distribution (Bréon and Goloub, 1998). The Rainbow Camera is an airborne imaging polarimeter developed at the NASA Goddard Space Flight Center to derive effective radii and widths of cloud droplet size distributions. The instrument has a 60° FOV and observes at 470, 550, 660, 760, 870, and 910 nm. It has flown on a research aircraft over the Amazon Basin (Correia et al., 2007).

Polarimetric remote sensing from space

Polarimetric observations of the Earth from orbit were acquired using a pair of Hasselblad cameras aboard NASA's Space Shuttle (Coulson et al., 1986). The cameras were outfitted with polarizing filters in perpendicular orientations, providing measurements of Q and I . Images were initially recorded on black and white film; color film was added on later flights. Egan et al. (1991) analyzed imagery over the Hawaiian islands and interpreted DOLP patterns as indications of soil texture, sea surface state, the presence of haze, and cloud optical depth.

Several spaceborne sensors designed for atmospheric trace gas measurements – the Global Ozone Mapping Experiment (GOME), the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY), and GOME-2 – make use of grating spectrometers that impart polarization sensitivity to the observations which, if uncorrected, lead to radiometric errors. To compensate, polarization measuring devices (PMDs) are incorporated to measure Q (SCIAMACHY also has sensitivity to U). The PMDs have been applied to atmospheric and surface remote sensing (e.g., Krijger et al., 2005; Lotz et al., 2009). GOME (Burrows et al., 1999) was launched aboard the European Space Agency (ESA) European Remote Sensing satellite ERS-2 in 1995, SCIAMACHY (Bovensmann et al., 1999) was launched aboard ESA's Envisat in 2002, and GOME-2 (Callies et al., 2003) was launched aboard ESA's MetOp-A in 2006.

The satellite Polarization and Directionality of the Earth's Reflectances (POLDER) instrument (Deschamps et al., 1994), developed by the French Centre National d'Etudes Spatiales (CNES), was specifically designed to obtain quantitative global, multiangle polarization imagery at a spatial resolution of ~ 6 km. POLDER was launched aboard the Japanese Advanced Earth Observing Satellite (ADEOS) in 1996 and acquired 8 months of data. A follow-on instrument, POLDER-2, was launched on ADEOS-2 in 2002 and acquired data for nearly a year. Both missions were cut short by spacecraft failures. A third instrument was launched in 2004 on the French PARASOL satellite. POLDER uses a mechanically stepped rotating filter/polarizer wheel, a fisheye refractive lens, and an area array CCD. Spectral bands are centered at 443, 490, 565, 670, 763, 765, 865, and 910 nm. Data at 443, 670, and 865 nm are acquired in three polarizer orientations to obtain I , Q , and U . DOLP uncertainty is ~ 1 –2% (Hagolle et al., 1999). A successor to POLDER, the

Multi-polarization Multidirectional Multispectral Instrument (3MI), extends the spectral coverage into the short-wave-infrared (Riedi, 2011). Launch of 3MI aboard Europe's EUMETSAT Polar System-Second Generation (EPS-SG) satellite system is planned for the 2020 time frame.

The Aerosol Polarimetry Sensor (APS) instrument, built by Raytheon as a successor to the airborne RSP, was intended to fly aboard NASA's Glory mission (Mishchenko et al., 2007), but due to a launch failure in March 2011, the spacecraft failed to reach orbit. The APS was designed to scan the FOVs of six bore-sighted telescopes in the along-track direction to acquire data at 410, 443, 555, 672, 865, 910, 1,378, 1,610, and 2,250 nm with a spatial resolution at nadir of ~ 6 km. Three of the six telescopes measure polarization at 0° and 90° and the other three measure at 45° and 135° (Peralta et al., 2007). As in RSP, the scan assembly design uses a pair of polarization-compensating mirrors. APS was to have provided the first highly accurate (non-imaging) multiangle polarimetric measurements from Earth orbit, with a DOLP uncertainty of $\sim 0.2\%$.

Conclusions

A variety of polarimeters have been deployed to view the Earth in reflected sunlight from suborbital and orbital vantage points. (Others have also flown on spacecraft missions to other planets). An earlier tabulation has been published by Coffeen (1974). Many investigations have demonstrated the value of polarimetry for characterizing terrestrial aerosols (e.g., Chowdhary et al., 2002; Herman et al., 2005; Waquet et al., 2009), clouds (e.g., Chepfer et al., 2001; Parol et al., 2004), and vegetation (e.g., Perry et al., 1997). Over the past several decades, dramatic improvements in polarimetric accuracy have resulted from increasingly sophisticated optical designs, the use of digital electronic detectors, and careful instrument calibration. Various imaging architectures have been developed (see, e.g., Tyo et al., 2006). The interested reader is referred to the vast body of literature that exists on the diverse scientific applications of polarimetry to the terrestrial environment.

Acknowledgment

The research to prepare this contribution was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Bibliography

- Auriol, F., Léon, J.-F., Balois, J.-Y., Verwaerde, C., François, P., Riedi, J., Parol, F., Waquet, F., Tanré, D., and Goloub, P., 2008. Multidirectional visible and shortwave infrared polarimeter for atmospheric aerosol and cloud observation: OSIRIS (Observing System Including Polarisation in the Solar Infrared Spectrum). *Proceedings of SPIE*, **7149**, 12.
- Barter, J. D., Thompson, H. R., Jr., and Richardson, C. L., 2003. Visible-regime polarimetric imager: a fully polarimetric, real-time imaging system. *Applied Optics*, **42**, 1620–1628.
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H., 1999. SCIAMACHY: mission objectives and measurement modes. *Journal of the Atmospheric Sciences*, **56**, 127–150.
- Bréon, F.-M., and Goloub, P., 1998. Cloud droplet effective radius from spaceborne polarization measurements. *Geophysical Research Letters*, **25**, 1879–1882.
- Burrows, J. P., Weber, M., Buchwitz, M., Razonov, V., Ladstätter, A., Richter, A., De Beer, R., Hoogen, R., Bramstedt, D., Eichmann, K. U., Eisenger, M., and Perner, D., 1999. The global ozone monitoring experiment (GOME): mission concept and first scientific results. *Journal of the Atmospheric Sciences*, **56**, 151–175.
- Cairns, B., Russell, E. E., LaVeigne, J. D., and Tennant, P. M. W., 2003. Research scanning polarimeter and airborne usage for remote sensing of aerosols. *Proceedings of SPIE*, **5158**, 33–44.
- Callies, J., Corpaccioli, E., Eisinger, M., Lefebvre, A., Munro, R., Perez-Albinana, A., Ricciarelli, B., Calamai, L., Gironi, G., Veratti, R., Otter, G., Eschen, M., and van Riel, L., 2003. GOME-2 the ozone instrument on-board the European METOP satellites. *Proceedings of SPIE*, **5158**, 1–11.
- Chepfer, H., Goloub, P., Riedi, J., Haan, J. F. D., Hovenier, J. W., and Flamant, P. H., 2001. Ice crystal shapes in cirrus clouds derived from POLDER/ADEOS-1. *Journal of Geophysical Research*, **106**, 7955–7966.
- Chipman, R. A., 1994. Polarimetry. In Bass, M. (ed.), *Handbook of Optics*, 2nd edn. New York: McGraw-Hill. Devices, Measurements and Properties, Vol. II. Chapter 22, 450 pp.
- Chowdhary, J., Cairns, B., and Travis, L. D., 2002. Case studies of aerosol retrievals over the ocean from multiangle, multispectral photopolarimetric remote sensing data. *Journal of the Atmospheric Sciences*, **59**, 383–397.
- Coffeen, D. L., 1974. Optical polarimeters in space. In *Planets, Stars and Nebulae Studied with Photopolarimetry; Proceedings of Twenty-third Colloquium*. Arizona: University of Arizona Press, pp. 189–217.
- Coffeen, D. L., Hämeen-Anttila, J., and Toubhans, R. H., 1975. Airborne infrared polarimeter. *Space Science Instrumentation*, **1**, 161–175.
- Correia, A. L., Fernandez-Borda, R., and Martins, J. V., 2007. Preliminary results of the Cloud-Aerosol Interaction Measurements (CLAIM) 2007 campaign on the Amazon Basin, Brazil. *Eos Transactions AGU*, **88**(52): Fall Meet. Suppl., Abstract A51D-0728.
- Coulson, K. L., Whitehead, V. S., and Campbell, C., 1986. Polarized views of the Earth from orbital altitude. *Proceedings of SPIE*, **637**, 35–41.
- Deschamps, P.-Y., Bréon, F. M., Leroy, M., Podaïre, A., Bricaud, A., Buriez, J.-C., and Sèze, G., 1994. The POLDER mission: instrument characteristics and scientific objectives. *IEEE Transactions on Geoscience and Remote Sensing*, **32**, 598–615.
- Deuzé, J. L., Devaux, C., Herman, M., Santer, R., Balois, J. Y., Gonzalez, L., Lecomte, P., and Verwaerde, C., 1989. Photopolarimetric observations of aerosols and clouds from balloon. *Remote Sensing of Environment*, **29**, 93–109.
- Deuzé, J. L., Bréon, F. M., Deschamps, P. Y., Devaux, C., Herman, M., Podaïre, A., and Roujean, J. L., 1993. Analysis of the POLDER (POLarization and Directionality of Earth's Reflectances) airborne instrument observations over land surfaces. *Remote Sensing of Environment*, **45**, 137–154.
- Diner, D. J., Davis, A., Hancock, B., Gutt, G., Chipman, R. A., and Cairns, B., 2007. Dual photoelastic modulator-based polarimetric imaging concept for aerosol remote sensing. *Applied Optics*, **46**, 8428–8445.
- Diner, D. J., Davis, A., Hancock, B., Geier, S., Rheingans, B., Jovanovic, V., Bull, M., Rider, D. M., Chipman, R. A., Mahler, A., and McClain, S. C., 2010. First results from a dual photoelastic-modulator-based polarimetric camera. *Applied Optics*, **49**, 2929–2946.

- Diner, D. J., Pingree, P. J., and Chipman, R. A., 2011. Novel airborne imaging polarimeter undergoes flight testing. *SPIE Newsroom*, 14 Nov 2011, doi:10.1117/2.1201111.003932.
- Egan, W.G., 1968. Aircraft polarimetric and photometric observations. In *Proceedings, Fifth International Symposium on Remote Sensing of Environment*. Ann Arbor: Environmental Research Institute of Michigan, pp. 169–189.
- Egan, W. G., Johnson, W. R., and Whitehead, V. S., 1991. Terrestrial polarization imagery obtained from the space shuttle: characterization and interpretation. *Applied Optics*, **30**, 435–442.
- Fernald, F. G., Herman, B. M., and Curran, R. J., 1969. Some polarization measurements of the reflected sunlight from desert terrain near Tucson, Ariz. *Journal of Applied Meteorology*, **8**, 604–609.
- Fernandez-Borda, R., Waluschka, E., Pellicori, S., Martins, J. V., Ramos-Izquierdo, L., Cieslak, J. D., and Thompson, P., 2009. Evaluation of the polarization properties of a Philips-type prism for the construction of imaging polarimeters. *Proceedings of SPIE*, **7461**, 15.
- Gál, J., Horváth, G., Meyer-Rochow, V. B., and Wehner, R., 2001. Polarization patterns of the summer sky and its neutral points measured by full-sky imaging polarimetry in Finnish Lapland north of the Arctic Circle. *Proceedings of the Royal Society London Series A*, **457**, 1385–1399.
- Hagolle, O., Goloub, P., Deschamps, P. Y., Cosnefroy, H., Briottet, X., Bailleul, T., Nicolas, J. M., Parol, F., Lafrance, B., and Herman, M., 1999. Results of POLDER in-flight calibration. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 1550–1566.
- Halajian, J. and Hallock, H. 1972. Principles and techniques of polarimetric mapping. In *Proceedings, Eight International Symposium on Remote Sensing of Environment*. Ann Arbor: Environmental Research Institute of Michigan, pp. 523–540.
- Hansen, J.E., and Coffeen, D.L., 1974. Analysis of cloud polarization measurements. In *Proceedings of the AMS Conference on Cloud Physics*, October 21–24, Tucson, pp. 350–356.
- Hansen, J. E., and Hovenier, J. W., 1974. Interpretation of the polarization of Venus. *Journal of the Atmospheric Sciences*, **31**, 1137–1160.
- Hariharan, T. A., 1969. An airborne polarimeter for atmospheric radiation studies. *Journal of Scientific Instruments (Journal of Physics E)*, **2**, 10–12.
- Herman, M., Balois, J. Y., Gonzalez, L., Lecomte, P., Lenoble, J., Santer, R., and Verwaerde, C., 1986. Stratospheric aerosol observations from a balloon-borne polarimetric experiment. *Applied Optics*, **25**, 3573–3584.
- Herman, M., Deuzé, J.-L., Marchand, A., Roger, B., and Lallart, P., 2005. Aerosol remote sensing from POLDER/ADEOS over the ocean: improved retrieval using a nonspherical particle model. *Journal of Geophysical Research*, **110**, D10S02.
- Hildum, E. A., and Spinhrne, J. D., 1992. An airborne system for collecting polarization imagery. *Proceedings of SPIE*, **1747**, 200–204.
- Hooper, B. A., Baxter, B., Piotrowski, C., Williams, J. Z., and Dugan, J., 2009. An airborne imaging multispectral polarimeter (AROSS-MSP). *OCEANS 2009, MTS/IEEE Biloxi-Marine Technology for our Future: Global and Local Challenges*, Biloxi, MS, 26–29 October 2009.
- Horváth, G., Bernáth, B., Suhai, B., Barta, A., and Wehner, R., 2002. First observation of the fourth neutral polarization point in the atmosphere. *Journal of the Optical Society of America A*, **19**, 2085–2099.
- Jensen, G. L., and Peterson, J. Q., 1998. Hyperspectral imaging polarimeter in the infrared. *Proceedings of SPIE*, **3437**, 42–51.
- Jones, S. H., Iannarilli, F. J., Hostetler, C., Cairns, B., Cook, A., Hair, J., Harper, D., Hu, Y., and Flittner, D., 2006. Preliminary airborne measurement results from the hyperspectral polarimeter for aerosol retrievals (HySPAR). In *NASA Earth Science Technology Conference Proceedings*, <http://esto.nasa.gov/conferences/ESTC2006/papers/b2p2.pdf>.
- Krijger, J. M., Aben, I., and Schrijver, H., 2005. Distinction between clouds and ice/snow covered surfaces in the identification of cloud-free observations using SCIAMACHY PMDs. *Atmospheric Chemistry and Physics*, **5**, 2729–2738.
- Lawrence, G. M., Thomas, G. E., Kohnert, R. A., and Westfall, J., 1994. An ultraviolet imaging polarimeter for observing polar mesospheric clouds. *Proceedings of SPIE*, **2266**, 232–241.
- Lotz, W. A., Vountas, M., Dinter, T., and Burrows, J. P., 2009. Cloud and surface classification using SCIAMACHY polarization measurement devices. *Atmospheric Chemistry and Physics*, **9**, 1279–1288.
- Millard, J. P., and Arvesen, J. C., 1972. Airborne optical detection of oil on water. *Applied Optics*, **11**, 102–107.
- Mishchenko, M. I., Cairns, B., Kopp, G., Schueler, C. F., Fafaul, B. A., Hansen, J. E., Hooker, R. J., Itchkawich, T., Maring, H. B., and Travis, L. D., 2007. Accurate monitoring of terrestrial aerosols and total solar irradiance: introducing the Glory Mission. *Bulletin of the American Meteorological Society*, **88**, 677–691.
- Parol, F., Buriez, J. C., Vanbauce, C., Riedi, J., Labonnote, L. C., Doutriaux-Boucher, M., Vesperini, M., Sèze, G., Couvert, P., Viollier, M., and Bréon, F.-M., 2004. Capabilities of multi-angle polarization cloud measurements from satellite: POLDER results. *Advances in Space Research*, **33**, 1080–1088.
- Peralta, R. J., Nardell, C., Cairns, B., Russell, E. E., Travis, L. D., Mishchenko, M. I., Fafaul, B. A., and Hooker, R. J., 2007. Aerosol polarimetry sensor for the Glory mission. *Proceedings of SPIE*, **6786**, 17.
- Perry, G. L., Stearn, J. A., Vanderbilt, V. C., Ustin, S. L., Diaz Barrios, M. C., Morrissey, L. A., Livingston, G. P., Bréon, F.-M., Bouffies, S., Leroy, M. M., Herman, M., and Balois, J.-Y., 1997. Remote sensing of high-latitude wetlands using polarized wide-angle imagery. *Proceedings-Spie The International Society For Optical Engineering*, **3121**, 375–386.
- Prosch, T., Hennings, D., and Raschke, E., 1983. Video polarimetry: a new imaging technique in atmospheric science. *Applied Optics*, **22**, 1360–1363.
- Rao, C. R. N., 1969. Balloon measurements of the polarization of light diffusely reflected by the Earth's atmosphere. *Planetary and Space Science*, **17**, 1307–1309.
- Rao, C. R. N., and Sekera, Z., 1967. A research program aimed at high altitude balloon-borne measurements of radiation emerging from the Earth's atmosphere. *Applied Optics*, **6**, 221–225.
- Riedi, J., 2011. 3MI: Operational monitoring of aerosols from EPS-SG. *Workshop on Observations and Modeling of Aerosol and Cloud Properties for Climate Studies*, Paris, France, 12–14 September 2011. <http://ebookbrowse.com/riedi-3mi-on-posteps-workshopparis-sept2011-pdf-d292740300>.
- Ruhtz, T., Pruesker, R., Hollstein, A., Bismarck, J. V., Starace, M., and Fischer, J., 2009. URMS/AMSSP (Universal Radiation Measurement System/Airborne Multi-Spectral Sunphoto- and Polarimeter). *Proceedings of SPIE*, **7475**, 8.
- Stephens, G. L., McCoy, R. F., Jr., McCoy, R. B., Gabriel, P., Partain, P. T., and Miller, S. D., 2000. A multipurpose scanning spectral polarimeter (SSP): instrument description and sample results. *Journal of Atmospheric and Oceanic Technology*, **17**, 616–627.
- Tozer, W. F., and Beeson, D. E., 1974. Optical model of noctilucent clouds based on polarimetric measurements from two sounding rocket campaigns. *Journal of Geophysical Research*, **79**, 5607–5612.
- Tyo, J. S., Goldstein, D. L., Chenault, D. B., and Shaw, J. A., 2006. Review of passive imaging polarimetry for remote sensing applications. *Applied Optics*, **45**, 5453–5469.

- Waquet, F., Léon, J.-F., Goloub, P., Pelon, J., Tanré, D., and Deuzé, J.-L., 2005. Maritime and dust aerosol retrieval from polarized and multispectral active and passive sensors. *Journal of Geophysical Research*, **110**, D10S10, doi:10.1029/2004JD004839.
- Waquet, F., Cairns, B., Knobelspiesse, K., Chowdhary, J., Travis, L. D., Schmid, B., and Mishchenko, M. I., 2009. Polarimetric remote sensing of aerosols over land. *Journal of Geophysical Research*, **114**, D01206.
- Witt, G., Dye, J. E., and Wilhelm, N., 1976. Rocket-borne measurements of scattered sunlight in the mesosphere. *Journal of Atmospheric and Solar-Terrestrial Physics*, **38**, 223–238.

Cross-references

[Aerosols](#)
[Cloud Properties](#)
[Optical/Infrared, Scattering by Aerosols and Hydrometeors](#)
[Reflected Solar Radiation Sensors, Multiangle Imaging](#)

REFLECTOR ANTENNAS

Yahya Rahmat-Samii
 Department of Electrical Engineering, University of
 California at Los Angeles, Los Angeles, CA, USA

Definition

Reflector antenna. An antenna consisting of one or more reflecting surfaces and a feed system for transmitting and/or receiving electromagnetic waves.

Introduction

Modern utilizations of reflector antennas in radars, satellite communications and tracking, remote sensing, radio astronomy, and the like have resulted in the development both of sophisticated reflector antenna configurations and of analytical and experimental design techniques. Reflector antennas typically take various configurations, some of the most popular ones being singly and doubly curved, single and multiple reflectors, and many others. The nature of applications and the resulting radiation patterns, such as pencil beam, contour beam, multifrequency, polarizations, and so on, dictate the actual configuration of reflector antennas.

One may classify reflector antennas in a variety of ways. Reflector antennas may be identified according to the types of radiation pattern, reflector surface, and feed. Pencil-beam reflectors are perhaps the most popular ones and are commonly used in point-to-point microwave communications and telemetry, since their patterns yield the maximum boresight gain and typically their beam directions are fixed at the time of antenna installation. In some applications, such as satellite systems, the uplink beam of these pencil-beam reflectors may be either fully steerable by reflector movements or capable of limited steering. Recent generations of satellite reflectors have produced other popular types of pattern classifications: contour (shaped) beams and multiple beams.

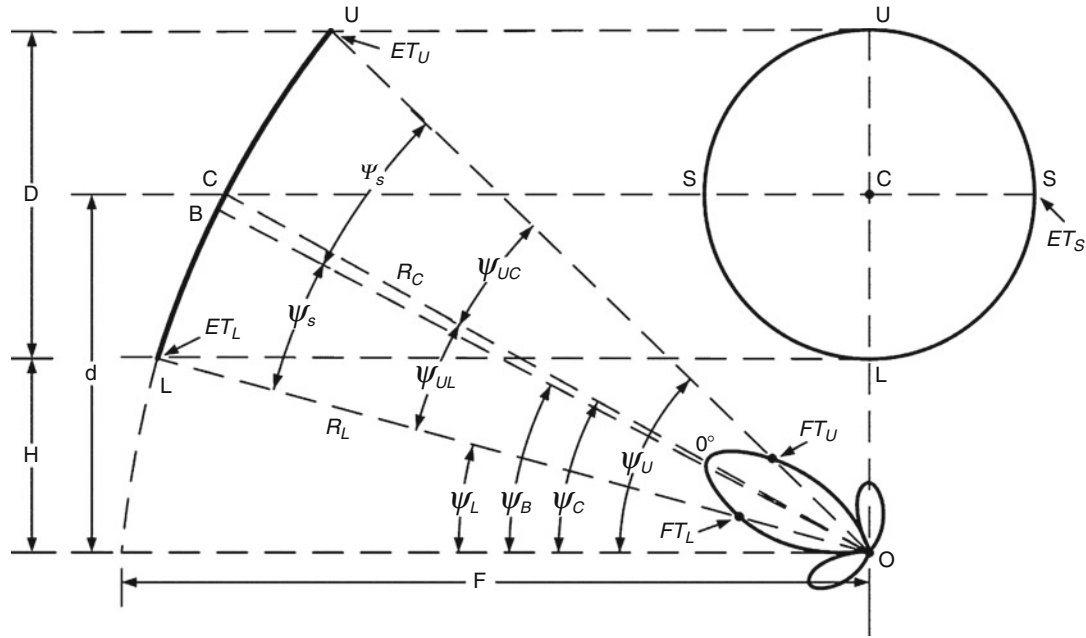
These applications require reflectors with improved off-axis beam characteristics, which result in more sophisticated configurations. There are many microwave communication antennas operating with one sense of polarization at a given frequency and require only reasonable discrimination between orthogonal polarizations. However, the current generation of microwave communication and remote sensing antennas operate with dual polarizations at the same frequency to enhance their frequency reuse capabilities or radar polarimetric properties. These requirements necessitate improved polarization performance and could be used as factors in pattern classification.

The objective of this entry is to briefly introduce reflector antennas by providing a summarized description of the performances of many representative reflector configurations. There are a few recently published handbooks, book chapters, and collections of papers that the reader is strongly advised to study (Silver, 1949; Jasik, 1961; Collin and Zucker, 1969; Skolnik, 1970; Love, 1976, 1978; Clarricoats and Poulton, 1977; Rudge and Adatia, 1978; Rudge et al., 1982; Rahmat-Samii, 1988, 2007).

Diffraction analysis techniques for reflector antennas

Advances in analytical and numerical techniques have allowed reflector antenna engineers to extensively utilize commercial or customized computer programs in the performance evaluation and design of reflector antennas. Accuracy of these techniques has been successfully checked against measured data.

Reflector diffraction analysis techniques may be classified as aperture field (AF) method (Collin and Zucker, 1969), geometrical optics (GO) and physical optics (PO) (Bucci et al., 1980; Rahmat-Samii and Galindo-Israel, 1980; Rahmat-Samii et al., 1981; Rahmat-Samii and Chueng, 1987), geometrical theory of diffraction (GTD) (Keller, 1962; Lee et al., 1979), method of moments (MoM), Gaussian beam method, or any hybridization of these. Clearly, there are advantages and disadvantages when comparing these techniques (Rahmat-Samii et al., 1991). Comparisons can be made based on the particular reflector configuration, far-field pattern domain, polarization, computation time, accuracy, and so on. MoM gives the most accurate result but it is impossible to use it economically for reflectors larger than many wavelengths. The aperture field method is not very accurate for offset configuration with displaced feeds when the edge diffracted rays are not included in the construction of the aperture fields. The GTD method is not easily applied in the caustics regions of pencil-beam reflectors. The PO method can take an excessive amount of computation time for large reflectors, in particular in the dual-reflector configurations, etc. Many advances have been reported to improve both the accuracy and the computation time. One of the most appealing techniques has been the application of GTD for the subreflector and PO for the main



Reflector Antennas, Figure 1 Geometrical parameters of an offset parabolic reflector (For symmetric cases, $H = -D/2$).

reflector, in conjunction with efficient expansions, such as Jacobi-Bessel expansion and sampling theorem.

In many applications the reflector surface may not be simply a solid PEC surface. For example, mesh reflectors are used in large unfurlable antennas, FSS (frequency selective surfaces) are used as subreflectors for frequency separation of feeds, and the reflector surface may be affected by snow. In all these cases, one is able to modify the PO integral and incorporate the effects of these surfaces (Rahmat-Samii and Lee, 1985; Imbriale et al., 1991; Rahmat-Samii and Tulinseff, 1993; Ip and Rahmat-Samii, 1998; Miura and Rahmat-Samii, 2005, 2007).

Offset (symmetric) parabolic reflectors

Parabolic-reflector antennas are still one of the most popular and practical reflector antenna configurations. The radiation characteristics of offset (symmetric) parabolic-reflector antennas illuminated by a single feed element are briefly discussed in this section. As is typical in any reflector design, there are too many almost-independent parameters which may be varied to achieve a particular design goal. To design an offset parabolic reflector, one must study the effects of such parameters as offset angles, illumination tapers, F/D ratios, locations and orientations of the feed, polarizations, etc., on such far-field pattern characteristics as scan loss, beamwidth, sidelobe level, cross-polarization level, and efficiency. Clearly it is not possible to perform an extensive study of all these parameters; rather, attempts are typically made to present the most important reflector characteristics based on a few key parameters. The results presented here can therefore be used

as a guideline for an initial design, which can then be refined by using computer programs and measured data.

Geometrical parameters

The geometry of an offset parabolic reflector with focal length F , diameter D , and offset height H is shown in Figure 1. In designing offset reflectors, any combination of the geometrical parameters may be given and the others may be constructed. Typically parameters D , F/D , and H/D are used as independent parameters and the others are obtained. The reader may want to refer to Rahmat-Samii (2007) for representative design curves.

Edge and feed tapers

The edge taper (ET) is a widely used parameter in characterizing the effects of the feed pattern on the far-field pattern of the reflector. The ratio of the field intensity at the reflector edge to the intensity at its center in decibels defines the edge taper. Although this definition is unambiguous when it is applied to symmetric reflectors, it can become ambiguous for offset reflectors. Customarily the feed taper (FT) is defined to overcome the ambiguity.

As observed in Figure 1, the feed taper FT_U in the upper-tip direction is defined as

$$FT_U = 20 \log \left[\frac{C(\psi_U - \psi_C)}{C(0^\circ)} \right]$$

where 0° defines the central direction (i.e., OC in Figure 1) and C is the feed pattern. Corresponding definitions for FT_L and FT_S can also be made at the lower and side angles

by using $\psi_C - \psi_L$ and ψ_S , respectively. The edge taper ET_U at the upper tip may now be expressed as

$$ET_U = FT_U + 20 \log(R_C/R_U)$$

where R_C and R_U are the path lengths from the feed to the center and the upper tip of the reflector, respectively. The second term $20 \log(R_C/R_U)$ is also called the path loss. Similar definitions can also be made for ET_L and ET_S . Note that for most cases of interest ET_U , ET_L , and ET_S have nearly equal values, and especially for symmetric reflectors, ET and FT only need to be defined at one edge (tip). Since ET directly controls the reflector's aperture amplitude taper, it has a more dominant effect on the far-field pattern than FT does. It is worth mentioning that it is not only the taper level which controls the reflector pattern but also the overall shape of the illumination distribution. In particular, the slopes of the illumination pattern at the reflector's edge can affect the sidelobe levels.

Radiation pattern characteristics

The following discussions on reflector pattern characteristics are based on the beams produced by on-focus feeds. The characteristics of the reflector antennas when the feed is displaced are investigated in Rahmat-Samii (2007). It is also assumed that the path losses are small (less than 0.5 dB) and, therefore, no substantial differences may be observed for path losses at different tips of the reflector. For these cases, $\psi_{UC} \approx \psi_{CL} \approx \psi_{SC} \approx \psi_S$ and $FT \approx ET$. The reader should attempt to properly interpret the results when the path losses are substantial.

In general, an offset parabolic reflector generates different far-field patterns in different cuts even when the on-focus feed has a symmetric pattern. Moreover, the pattern can be slightly asymmetric in the plane of offset depending on the F/D ratio (Rahmat-Samii, 1984). However, for the results presented here, a large F/D ratio is used in order to reduce the path loss effects and asymmetry of the pattern. Figure 2 shows the half-power beamwidth, first-null, second-null positions, and first sidelobe positions as functions of edge tapers. Note that the results shown here are for $\cos^q(\theta)$ -type feed amplitude patterns and the pattern characteristics depend heavily on the actual feed pattern description for edge tapers beyond -20 dB. Also shown in Figure 3 are the first sidelobe level, taper efficiency η_t , spillover efficiency η_s , and overall efficiency η as functions of edge tapers. It is readily observed that the resulting overall efficiency $\eta = \eta_s \eta_t$ is maximized for edge tapers about -11 dB with a value of 81 %. Some representative far-field patterns for different edge tapers are plotted in Figure 4. It can also be observed that for edge tapers in the neighborhood of -20 dB, the first sidelobe starts to merge with the main beam, which results in a widened beam. Again the exact distribution of the feed pattern can have a significant effect on the pattern characteristics for this level of edge taper.

The cross-polarization level of the radiated far-field is another performance parameter of interest. This topic has

been addressed in Love (1978), Chu and Turrin (1973), and Ludwig (1973) for both symmetric and offset reflectors, and some representative cases are presented in this entry. Figure 5 shows some typical patterns of cross-polarized field of a symmetric reflector fed by an unbalanced linearly polarized feed (a feed with different E- and H-plane patterns). The maximum of cross-polarized fields occurs in the plane $\phi = 45^\circ$, and the level of the cross-polarized field decreases substantially as the feed becomes more balanced. Note that these results are dependent on the F/D ratio and the edge taper.

In contrast to symmetric reflectors with balanced and linearly polarized feeds, which produce very low levels of cross-polarized fields, offset parabolic reflectors can generate high level of cross-polarization, depending on the tilt angle of the feed axis with respect to the reflector axis for linearly polarized feeds. For instance, Figure 6 presents the cross-polarized fields for various feed-axis tilt angles for a fixed offset reflector configuration, where the feed pattern is chosen to be isotropic in order to better illustrate the generation of the cross-polarized fields. Clearly, a high level of cross-polarized field should be expected in practice, since the feed axis is always tilted toward the center of the offset reflector in order to reduce spillover. The levels of cross-polarized fields for different values of bisect angle ψ_B and the half-subtended angle ψ_S (see Figure 1) can be found in Chu and Turrin (1973). It is worth mentioning that the cross-polarized field is predominantly observed in the plane $\phi = 90^\circ$ (normal to the plane of the offset) for offset parabolic reflectors.

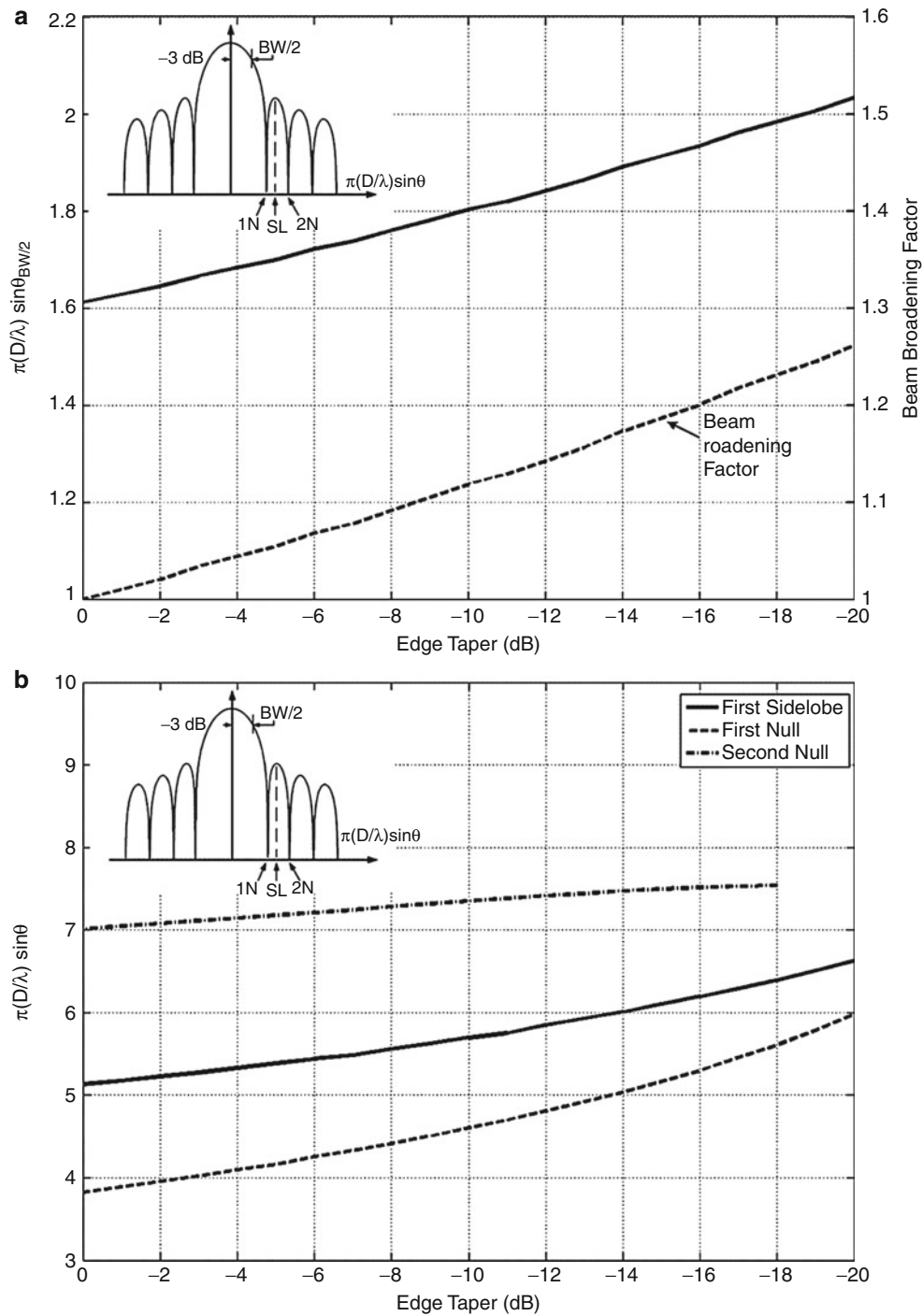
When illuminated with balanced circularly polarized feeds, both symmetric and offset parabolic reflectors generate very low levels of cross-polarized fields. However, an additional feature, which is referred to as the beam squint, is observed for offset parabolic reflectors with circularly polarized feeds. In the plane normal to the plane of offset, the beam peak is shifted from the boresight by a small amount, which depends on the tilt angle of the feed axis and the reflector geometry. The following expression is a good approximation to estimate the beam squint:

$$\sin \theta_S = \mp \frac{\sin \psi_B}{4\pi(F/\lambda)}$$

where $\mp$ signs are for right and left circularly polarized beams, F is the focal length, and θ_S is the amount of squint. The squint angle obtained from diffraction analysis and experimental data (Chu and Turrin, 1973) agrees well with the approximate formula. A representative example is shown in Figure 7 (Chu and Turrin, 1973), where again ψ_B is used to characterize the reflector. Attempts have also been made to correct the undesired beam squint effect in certain reflector designs (Duan and Rahmat-Samii, 1991).

Other conventional reflector antennas

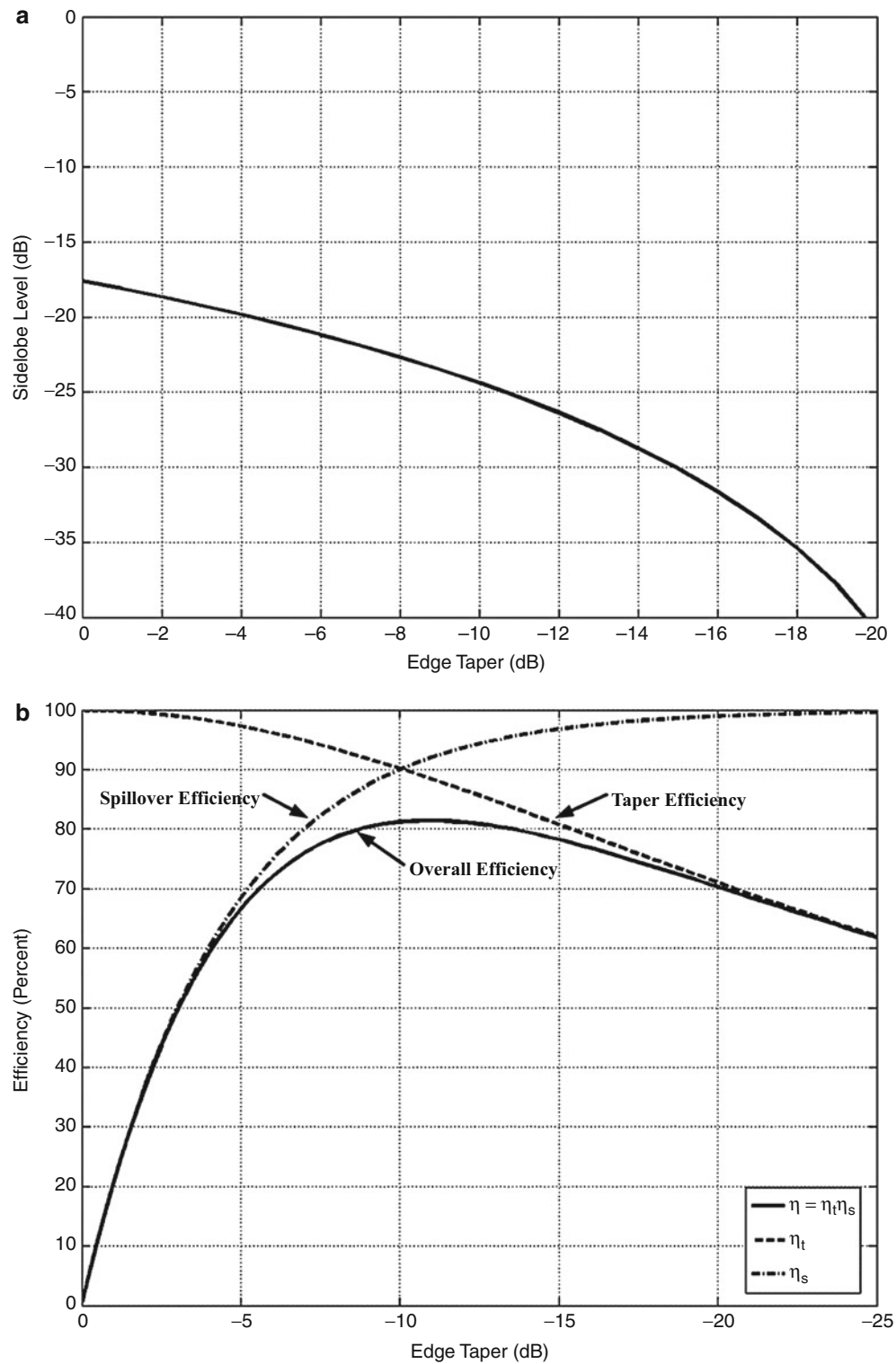
Different types of reflector antennas have been developed for a variety of applications. The parabolic arc can be used



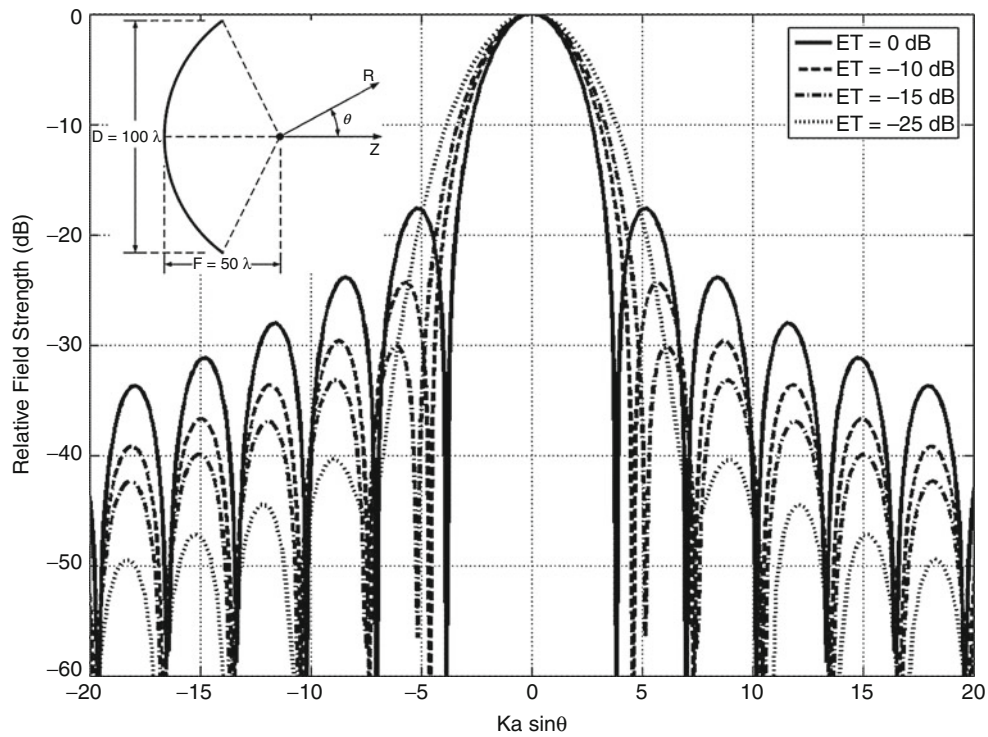
Reflector Antennas, Figure 2 (a) Half-power beamwidth and (b) first-null, second-null, and first sidelobe positions as functions of the edge taper ET , where D/λ is the diameter of the reflector in terms of wavelengths.

as a generating curve for a number of useful reflector surfaces. When moved so that the focal point travels along a straight line, the curve generates a parabolic cylinder. With a limited motion along this line, the reflector for a pillbox is generated. When the focal point is moved

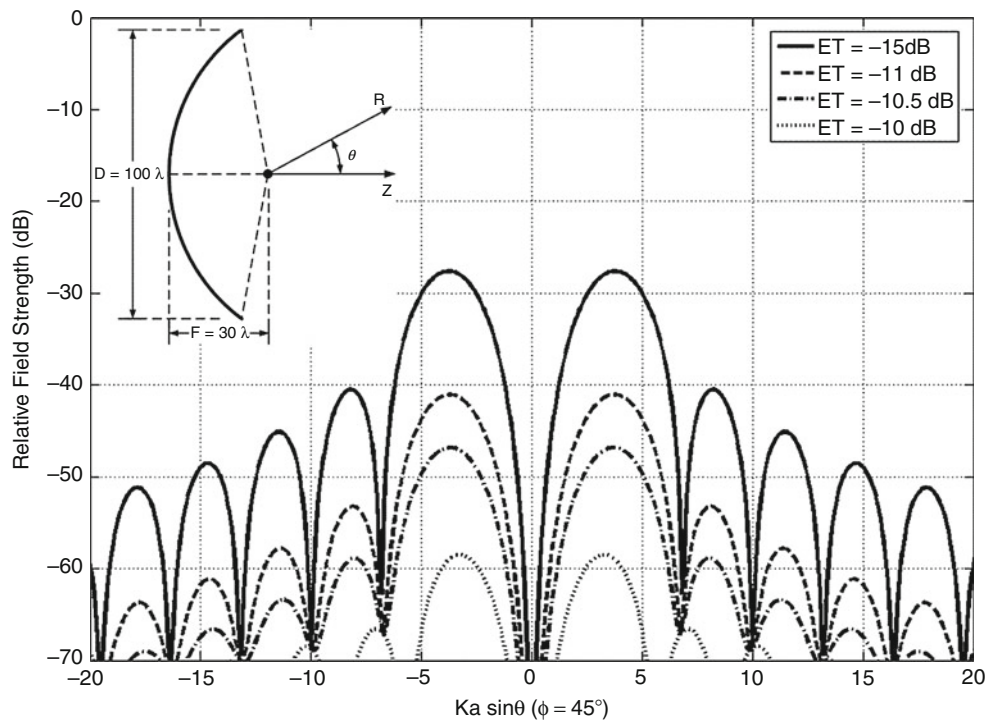
along a circle enclosing the arc, an hourglass surface is generated, and for a circle which does not enclose the arc, a parabolic torus is generated. This entry only discusses some representative designs due to the limited space.



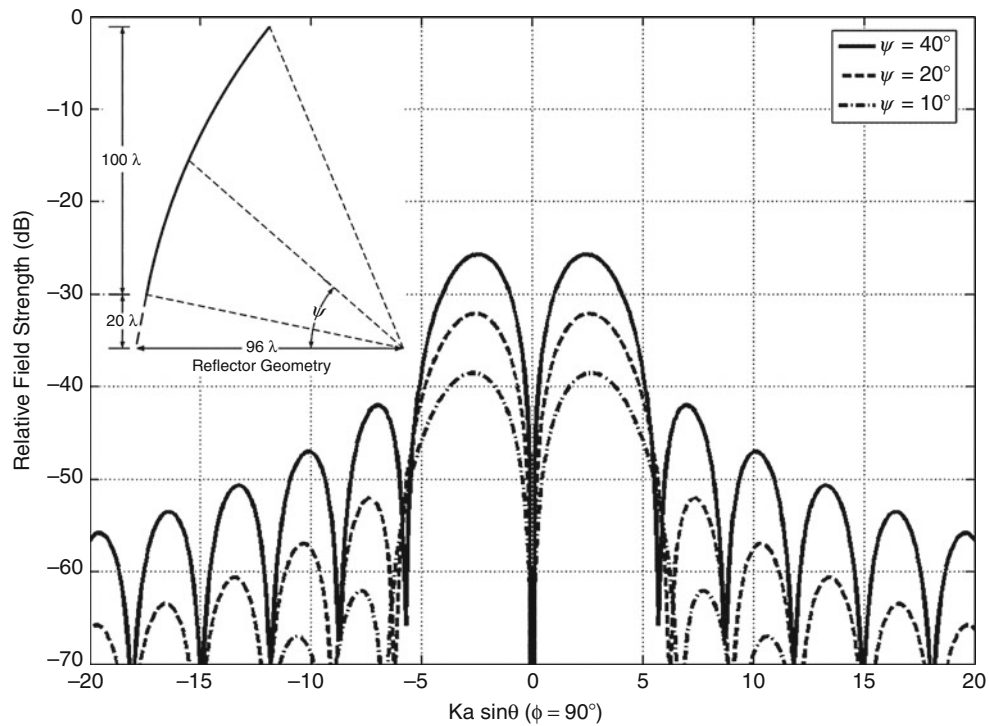
Reflector Antennas, Figure 3 (a) First sidelobe level and (b) taper efficiency, spillover efficiency, and overall efficiency as functions of edge tapers for $\cos^q(\theta)$ -type feed patterns.



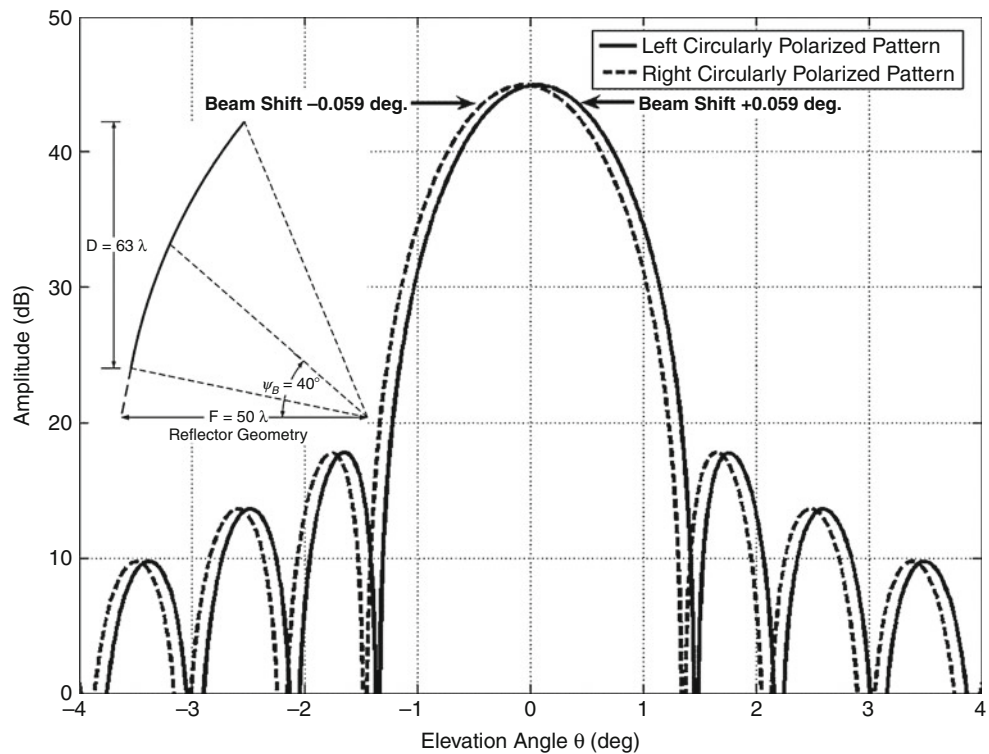
Reflector Antennas, Figure 4 Representative reflector normalized far-field patterns for different edge tapers.



Reflector Antennas, Figure 5 Cross-polarized field (normalized with respect to the peak of the co-polarized field) of a symmetric reflector ($F/D = 0.3$) illuminated by a linearly polarized feed with unbalanced E - and H -patterns.



Reflector Antennas, Figure 6 Cross-polarized field in the plane $\phi = 90^\circ$ (normal to the offset plane) for an offset parabolic reflector for various feed-axis tilt angles ψ and illuminated by a balanced linearly polarized feed.



Reflector Antennas, Figure 7 Squinted far-field patterns in the plane $\phi = 90^\circ$ of an offset parabolic reflector illuminated by right and left circularly polarized feed located at its focal point.

Parabolic cylinder

A parabolic cylinder is constructed by moving a parabolic arc along a straight line. Being a singly curved surface, problems of construction and maintaining tolerances are somewhat easier than those encountered in doubly curved surfaces such as the paraboloid. A line-source feed is usually employed to illuminate the parabolic cylinder, and its aperture coincides with the focal line of the cylinder. The line source may be formed in a number of ways. When the aperture of the line source contains energy in phase along the length, a cylindrical wave front is generated and further focused by the parabolic cylinder into a plane wave in the direction of the axis of the parabolic arc. If a linear-phase variation exists along the length, a conical wave front is resulted, which then produces a plane wave tilted from the parabolic axis by an angle equal to the half angle of the cone.

The parabolic cylinder can also be used as the reflector in the pillbox, or cheese, antenna. It is constructed by two parallel plates which cut through a parabolic cylinder perpendicular to the cylinder. Typically positioned in the center of the aperture, the feed blocks a significant portion of the open region. Large sidelobes in the pattern of the pillbox are resulted due to the feed blockage, and the feed backlobes can also significantly affect the reflector pattern. Furthermore, a portion of the feed's radiation is reflected back into itself to produce a standing wave. An improvement in performance can be achieved when a half pillbox is used in the same fashion as the offset parabolic cylinder. The arc of the parabola extends from the vertex to the 90° point. The feed horn is directed at the 45° point. Although the illumination is asymmetrical, good sidelobes are obtained. A recent spaceborne concept was designed and tested in Rahmat-Samii et al. (2005).

Parabolic torus

A parabolic torus reflector is formed by rotating the parabolic arc about a vertical axis positioned on the concave side of the arc. It can be generated by a symmetrical arc or by the more widely used offset arc. The entire reflector is illuminated by multiple adjacent feeds, each of which uses portions of the reflector surface. A relatively compact structure can be obtained with multiple usage of the surface. A recent spaceborne application using this concept was suggested in Hoferer and Rahmat-Samii (1998); Njoku et al. (1999).

Spherical reflector antennas

When a point source is positioned at the focus, caustic region, of a spherical reflector, the reflected wave front in the aperture plane is not planar, and the subsequent phase error is known as spherical aberration. Although not a perfect focusing device, spherical reflectors are capable of focusing paraxial rays quite well. The design of the sphere is based upon the phase-error formula. The error curve is used out to the point at which the phase error is zero. An argument can be made that some phase error

beyond this point can be accepted. In practical designs, two points on the phase-error curves, namely, the maximum value and the edge (zero) value which defines the maximum value of the aperture radius r , are used. A specific curve using proper geometrical parameters is selected so that a simple ratio, namely, maximum phase error divided by r , is satisfied. Therefore, a practical sphere is limited in extent in order to keep the phase error acceptable.

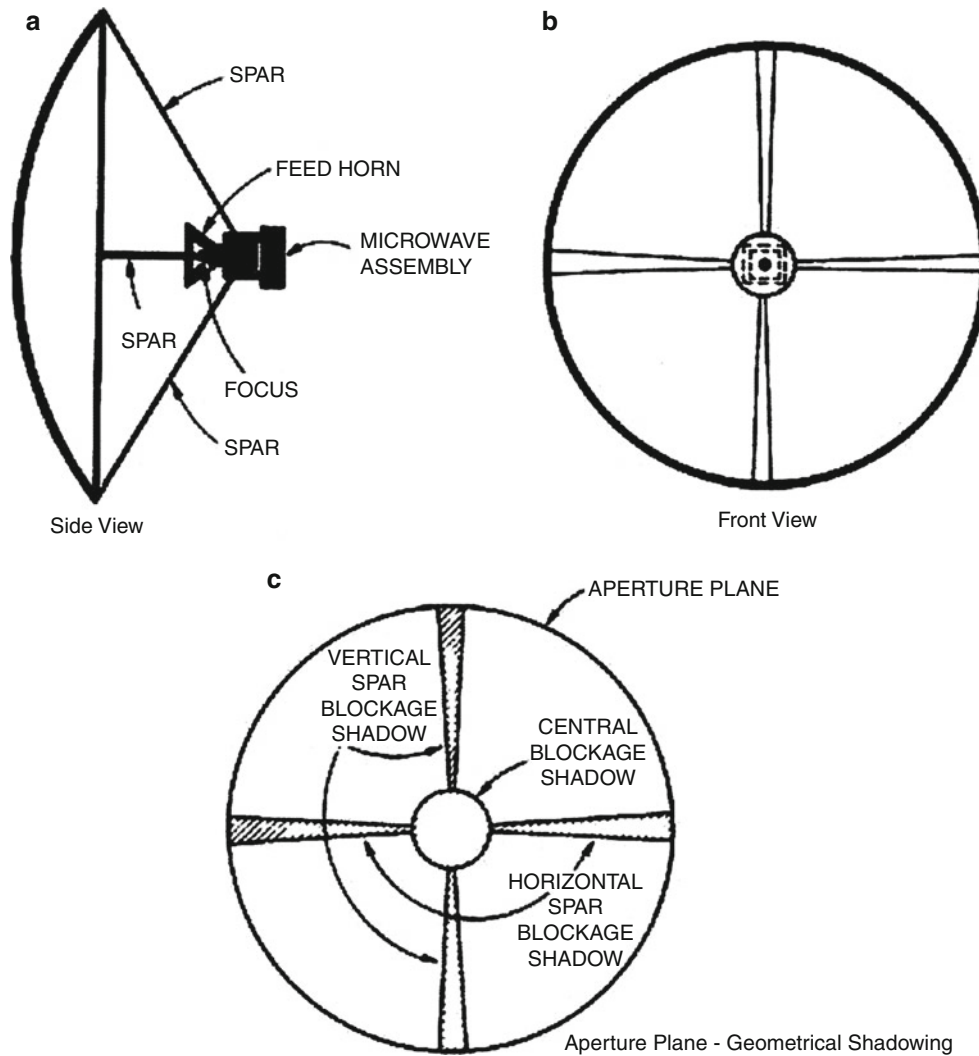
If the spherical reflector can have efficiency of the order of 10 %, it will give patterns which do not change with the beam moved off axis (Li, 1959). In this instance, the sphere is fed from a large feed horn which confines its radiation to a small portion of the surface. The reflector edges which were the problem in the normal sphere are then eliminated. In effect, this solution uses a very long-length system in which the paraboloid and the sphere become very much alike.

An auxiliary correcting apparatus is usually required if a spherical reflector with large-aperture size is of interest. The potential of using a large spherical reflector antenna as a hurricanes-monitoring radar from geostationary orbit was investigated in Bahadori and Rahmat-Samii (2005a) and Xu and Rahmat-Samii (2009). A spherical reflector operating at Ka band (35.6 GHz), with a 35.5 m aperture and 28 m effectively illuminated area, was chosen to meet the most challenging demands of producing an unconventionally high gain and large number of beams scanned up to 200 beamwidths. Using a properly designed planar array feed or a sub-reflector array, the spherical aberration can be effectively compensated.

Large-aperture reflector antennas

In the 1970s, reflector antennas with electrically large apertures, typically greater than 60 wavelengths, became relatively common. Dominantly analyzed by geometrical optics and often refined by physical optics (Rusch and Potter, 1970), these designs focus on some parameters of concern: gain, beamwidth, sidelobe level, polarization and cross-polarization, beam efficiency, antenna noise temperature, and, if more than one beam position is needed, the variation of these parameters with scan angle. The design of the feed or, in the case of more complex antenna systems, the feed system and subreflector profoundly affects the parameters through illumination control, feed polarization properties, and spillover past the subreflector or the main reflector. Also the structural design affects electrical performance through the scattering of energy from surface-panel irregularities and interpanel gaps, from subreflector and/or feed-support spars, and from the correlation of this scattering arising from the structural design and alignment procedures.

The gain and complexity of spaceborne antennas have steadily increased since the inception of satellite communications. A need to increase the effective isotropic radiated power (EIRP) in a power-limited environment demands higher gain, while other requirements such as



Reflector Antennas, Figure 8 Geometrical configuration of a front-fed parabolic-reflector antenna.

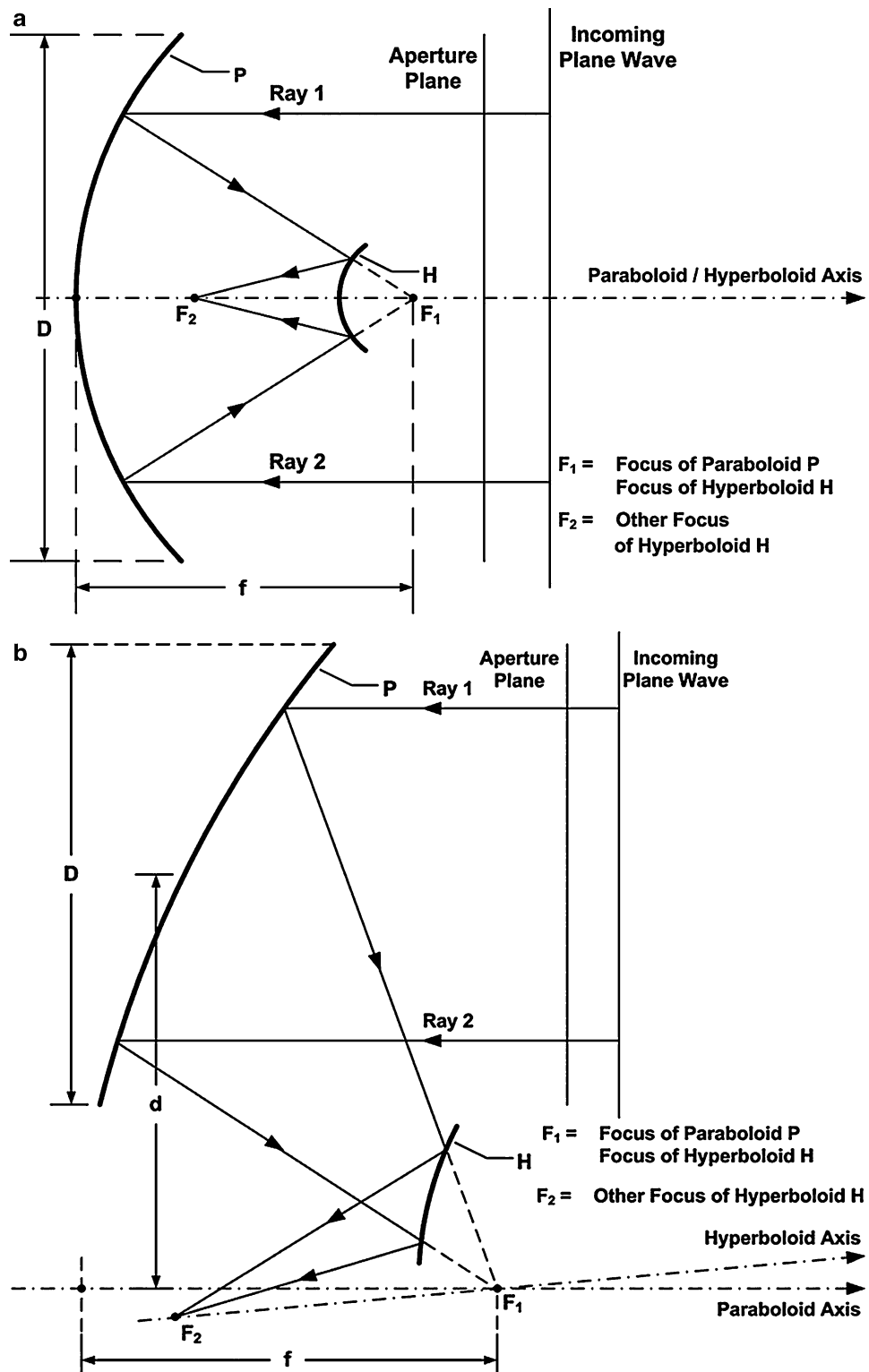
beam shaping or reconfigurability necessitate more complex antenna configurations. Electrically large reflectors are usually fed by simpler antennas such as horns or array feeds or by folded-optics systems involving subreflectors. In many instances, the antennas are designed to produce two coaxial beams with orthogonal polarizations over near-octave bandwidths. The feed systems may then terminate in complex microwave circuits involving polarizers and orthomode transducers to separate or combine orthogonally polarized beams as well as in microwave multiplexers to separate or combine different frequencies.

Reflector antennas with electrically large apertures have their principal applications in radars and radiometry, where their high gain and narrow beamwidths enhance radar range and angular resolution; in radio astronomy and other deep-space telemetry, where the same parameters enhance sensitivity and resolution of stellar radio

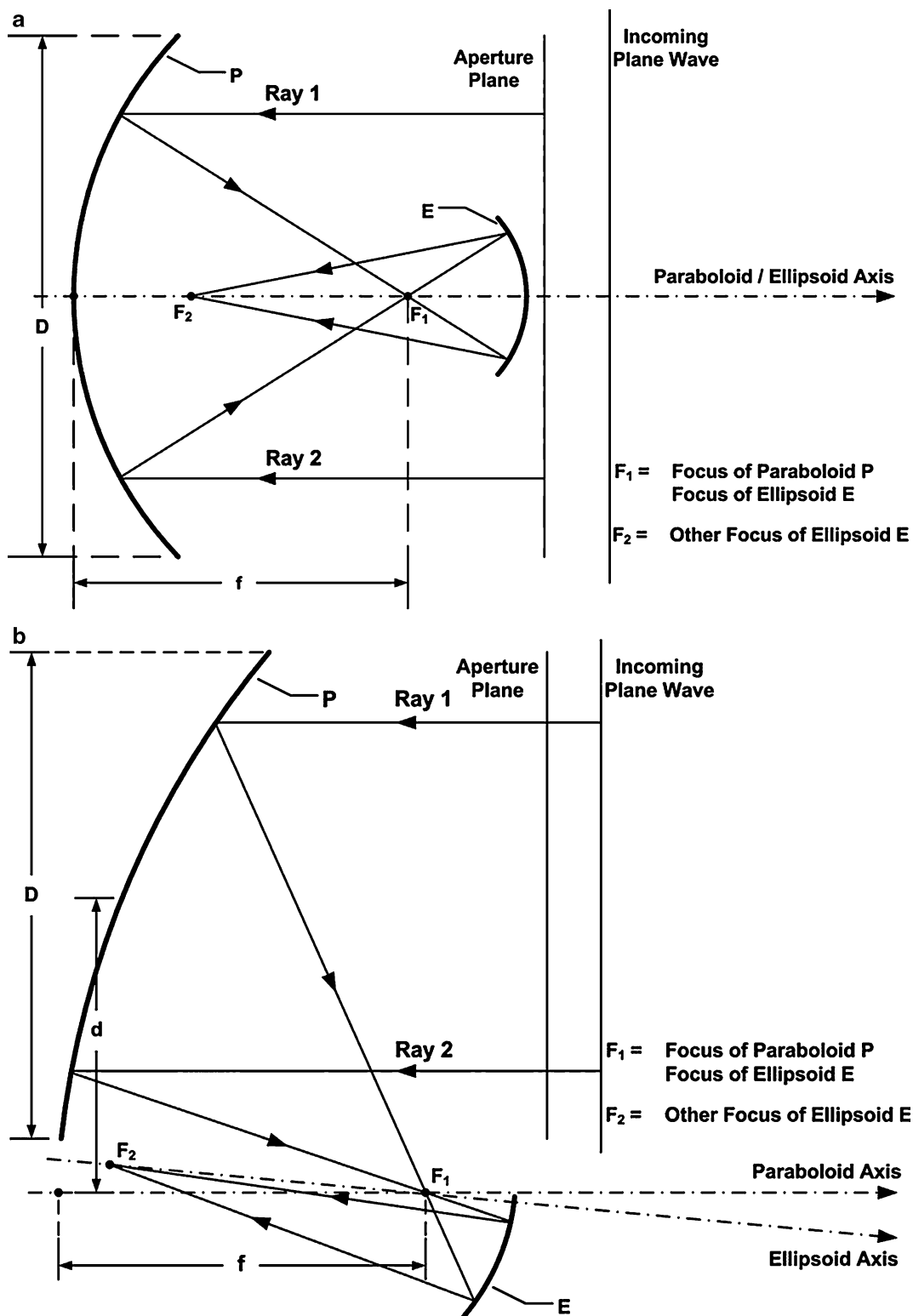
sources; and in microwave communications and remote sensing, where high gain, high beam efficiency, low noise, and sidelobe control enhance EIRP, receive sensitivity, and isolation from interference. For reflector antennas of diameter in the range of 60–1,000 wavelengths, gains range from 45 to 70 dBi, half-power beamwidths from 1.5° to 0.05° , and average wide-angle sidelobe levels down to -20 dBi (almost 90 dB below the peak of the main beam). Clearly, these features depend upon the feed system as well as the reflectors.

Folded optics: Cassegrain and Gregorian reflectors

The simple symmetric front-fed reflector has disadvantages due to the blockages of the aperture by the feed system and by the feed supports. Two views of a front-fed parabolic-reflector antenna with feed supports extending from the reflector edge are presented in Figure 8. It is



Reflector Antennas, Figure 9 Geometrical configuration of a Cassegrain antenna: (a) symmetric and (b) offset.



Reflector Antennas, Figure 10 Geometrical configuration of a Gregorian antenna: (a) symmetric and (b) offset.

observed that the feed and its supports intercept partial radiation energy reflected from the reflector. It is important to note that supports which are not on the edge intercept energy from the feed before it reaches the reflector. In a geometrical-optics sense, they cast a shadow upon the aperture plane, as shown in the third view in Figure 8. This shadow, called aperture blockage, reduces gain and affects near-in sidelobes, raising some and lowering others. Moreover, the blocked radiation is further scattered by the feed and its supports and can significantly increase wide-angle sidelobes. The front-fed reflector typically has an efficiency of 55–60 %. One possible approach to increase efficiency is to employ folded-optics designs.

The radiation energy may be reflected by one or more subreflectors in the path from the feed to pencil-beam reflectors. As illustrated in Figure 9, the Cassegrain reflector system folds the incoming-ray trajectories back in the direction of the vertex of the paraboloid P by reflection from the hyperboloidal subreflector H . The focusing properties of the system derive from the collocation of one focus of the hyperboloid H with that of the paraboloid at F_1 . An incoming plane wave from the direction of the paraboloid axis is reflected from the paraboloid and then from the hyperboloid and focuses at F_2 , the other focus of the hyperboloid H . Therefore, the incident energy can be received by the feed with its phase center at F_2 .

The geometry of the Gregorian reflector system is shown in Figure 10. The subreflector is an ellipsoid E with its near focus collocated at F_1 with the focus of the paraboloid. To create a confocal geometry, the phase center of the feed should then lie at F_2 , the other focus of the ellipsoid E .

For the same focal length, the Cassegrain system is more compact than the Gregorian system. Arguments have been made that Gregorian systems may provide improved wide-angle sidelobes in apertures of the order of 200 wavelengths. However, the sidelobes of these folded antennas are not especially low, because of blockages of the feed system, subreflector, and supports.

A number of sources contribute to the high sidelobes in a folded-optics system, which include (1) radiation-pattern sidelobes from the aperture illumination, (2) spillover and edge diffraction from the subreflector illuminated by the feed system, (3) spillover and edge diffraction from the main reflector illuminated from the subreflector, (4) central blockage (subreflector and feed system), (5) spar (subreflector support) blockage and scattering, (6) deviation of reflector surfaces from the desired contours, and (7) effects of gaps between panels (Williams, 1965; Zucker, 1968; Bao, 1969; Schrank, 1985). Some recent papers (Rahmat-Samii, 1985; Ling et al., 1986; Wu and Rahmat-Samii, 1991; Bahadori and Rahmat-Samii, 2005b) and a review paper (Corkish, 1990) on the effects of reflector surface distortions on sidelobe levels cover the various approaches to this subject in some detail and provide a wealth of references.

The increased interest in large, spaceborne antennas and in sidelobe control has given rise to increased interest

in and use of offset configurations, especially for multibeam and scanning applications (Rudge and Adatia, 1978; Lee and Rahmat-Samii, 1981). The conventional pencil-beam Cassegrain and Gregorian configurations have been studied to provide design techniques for the offset cases (Rusch et al., 1990). A technique for increasing antenna efficiency by extending the subreflector in an offset Cassegrain antenna has been analyzed (Rahmat-Samii, 1986). When offset Cassegrain or Gregorian reflector systems (Figures 9b and 10b) are used in beam-scanning or multibeam applications, feeds must be moved or located in the positions that provide optimal beam performance. This turns out to be a hyperboloid for the offset Cassegrain and an ellipsoid for the offset Gregorian (Rahmat-Samii and Galindo-Israel, 1981).

Advances in real-time control techniques and data processing have become more matured, allowing the possibility of correcting for distortions. Adaptive phased-array feeds can be used to compensate for deterioration of scanning (Rahmat-Samii, 1990). Schell's multiplate antenna (Schell, 1966) provides all the necessary characteristics for adaptive removal of both atmospheric and surface distortions. The advent of modern control electronics and drive mechanisms should make the multiplate antenna a candidate approach for large-aperture reflectors. More recently, a novel electronically reconfigurable subreflectarray using microelectromechanical systems (MEMS) technology was proposed to accomplish real-time reflector surface distortion compensation (Rajagopalan et al., 2008; Xu et al., 2009).

Summary

A reflector antenna consists of one or more reflecting surfaces and a feed system. Due to its excellent radiation performances, versatility, and manufacturing readiness, it has been widely used in radars, satellite communications, deep-space telemetry, remote sensing, and radio astronomy. Reflector antennas have various types and designs, among which parabolic reflectors are the most commonly used in practice. By selecting proper geometrical parameters and feed illumination, the reflector radiation characteristics such as gain, beamwidth, sidelobe level, and cross-polarization can be well tailored for different goals. Some more complex systems such as spherical reflectors and Cassegrain and Gregorian reflectors have also been developed for more advanced applications.

Bibliography

- Bahadori, K., and Rahmat-Samii, Y., 2005a. An array-compensated spherical reflector antenna for a very large number of scanned beams. *IEEE Transactions on Antennas and Propagation*, **53**, 3547–3555.
- Bahadori, K., and Rahmat-Samii, Y., 2005b. Characterization of effects of periodic and aperiodic surface distortions on membrane reflector antennas. *IEEE Transactions on Antennas and Propagation*, **53**, 2782–2791.
- Bao, V. T., 1969. Influence of correlation interval and illumination taper in antenna tolerance theory. *Proceedings of the IEE*, **116**, 195–202.

- Bucci, O. M., Franceschetti, G., and Elia, G. D., 1980. Fast analysis of large antennas – A new computational philosophy. *IEEE Transactions on Antennas and Propagation*, **28**, 306–310.
- Chu, T. S., and Turrin, R. H., 1973. Depolarization properties of offset reflector antennas. *IEEE Transactions on Antennas and Propagation*, **21**, 339–345.
- Claricoats, P. J. B., and Poulton, G. T., 1977. High-efficiency microwave reflector antennas – a review. *Proceedings of the IEEE*, **65**, 1470–1504.
- Collin, E. C., and Zucker, F. J. (eds.), 1969. *Antenna Theory, Parts 1 and 2*. New York: McGraw-Hill.
- Corkish, R. P., 1990. A survey of the effects of reflector surface distortions on sidelobe levels. *IEEE Antennas and Propagation Magazine*, **32**, 6–11.
- Duan, D. W., and Rahmat-Samii, Y., 1991. Beam squint determination in conic-section reflector antennas with circularly polarized feeds. *IEEE Transactions on Antennas and Propagation*, **39**, 612–619.
- Hoferer, R. A., and Rahmat-Samii, Y., 1998. RF characterization of an inflatable parabolic torus reflector antenna for space-borne applications. *IEEE Transactions on Antennas and Propagation*, **46**, 1449–1457.
- Imbriale, W. A., Galindo-Israel, V., and Rahmat-Samii, Y., 1991. On the reflectivity of complex mesh surfaces. *IEEE Transactions on Antennas and Propagation*, **39**, 1352–1365.
- Ip, A., and Rahmat-Samii, Y., 1998. Analysis and characterization of multilayered reflector antennas: rain/snow accumulation and deployable membrane. *IEEE Transactions on Antennas and Propagation*, **46**, 1593–1605.
- Jasik, H., 1961. *Antenna Engineering Handbook*. New York: McGraw-Hill.
- Keller, J. B., 1962. Geometrical theory of diffraction. *Journal of the Optical Society of America*, **52**, 116–130.
- Lee, S. W., and Rahmat-Samii, Y., 1981. Simple formulae for designing on offset multibeam parabolic reflector. *IEEE Transactions on Antennas and Propagation*, **29**, 472–478. corrections: 30:323.
- Lee, S. W., Cramer, P., Jr., Woo, K., and Rahmat-Samii, Y., 1979. Diffraction by an arbitrary subreflector: GTD solution. *IEEE Transactions on Antennas and Propagation*, **27**, 305–316.
- Li, T., 1959. A study of spherical reflectors as wide-angle scanning antennas. *IRE Transactions on Antennas and Propagation*, **7**, 223–226.
- Ling, H., Lo, Y. T., and Rahmat-Samii, Y., 1986. Reflector sidelobe degradation due to random surface errors. *IEEE Transactions on Antennas and Propagation*, **34**, 164–172.
- Love, A. W. (ed.), 1976. *Electromagnetic Horn Antennas*. New York: IEEE Press.
- Love, A. W. (ed.), 1978. *Reflector Antennas*. New York: IEEE Press.
- Ludwig, A. C., 1973. The definition of cross polarization. *IEEE Transactions on Antennas and Propagation*, **21**, 116–119.
- Miura, A., and Rahmat-Samii, Y., 2005. RF characteristics of spaceborne antenna mesh reflecting surfaces: application of periodic method of moments. *Microwave and Optical Technology Letters*, **47**, 365–370.
- Miura, A., and Rahmat-Samii, Y., 2007. Spaceborne mesh reflector antennas with complex weaves: extended PO/periodic-MoM analysis. *IEEE Transactions on Geoscience and Remote Sensing*, **55**, 1022–1029.
- Njoku, E. G., Rahmat-Samii, Y., Sercel, J., Wilson, W. J., and Moghaddam, M., 1999. Evaluation of an inflatable antenna concept for microwave sensing of soil moisture and ocean salinity. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, 63–78.
- Rahmat-Samii, Y., 1984. A comparison between GO/aperture field and physical optics methods for offset reflectors. *IEEE Transactions on Antennas and Propagation*, **32**, 301–306.
- Rahmat-Samii, Y., 1985. Effects of deterministic surface distortions on reflector antenna performance. *Annales des Telecommunications*, **40**, 350–360.
- Rahmat-Samii, Y., 1986. Subreflector extension for improved efficiencies in Cassegrain antennas: GTD/PO analysis. *IEEE Transactions on Antennas and Propagation*, **34**, 1266–1269.
- Rahmat-Samii, Y., 1988. Reflector antennas. In Lo, Y. T., and Lee, S. W. (eds.), *Antenna Handbook*. New York: Van Nostrand Reinhold.
- Rahmat-Samii, Y., 1990. Array feeds for reflector surface distortion compensation: concepts and implementation. *IEEE Antennas and Propagation Magazine*, **32**, 20–26.
- Rahmat-Samii, Y., 2007. Reflector antennas. In Volakis, J. (ed.), *Antenna Engineering Handbook*, 4th edn. New York: McGraw-Hill.
- Rahmat-Samii, Y., and Chueng, R., 1987. Nonuniform sampling techniques for antenna applications. *IEEE Transactions on Antennas and Propagation*, **35**, 268–279.
- Rahmat-Samii, Y., and Galindo-Israel, V., 1980. Shaped reflector antenna analysis using the Jacobi-Bessel series. *IEEE Transactions on Antennas and Propagation*, **28**, 425–435.
- Rahmat-Samii, Y., and Galindo-Israel, V., 1981. Scan Performance of dual offset reflector antennas for satellite communications. *Radio Science*, **16**, 1093–1099.
- Rahmat-Samii, Y., and Lee, S. W., 1985. Vector diffraction analysis of reflector antennas with mesh surfaces. *IEEE Transactions on Antennas and Propagation*, **33**, 76–90.
- Rahmat-Samii, Y., and Tulintseff, A. N., 1993. Diffraction analysis of frequency selective reflector antennas. *IEEE Transactions on Antennas and Propagation*, **41**, 476–487.
- Rahmat-Samii, Y., Mittra, R., and Galindo-Israel, V., 1981. Computation of Fresnel and Fraunhofer fields of planar apertures and reflector antennas by Jacobi-Bessel series – a review. *Journal of Electromagnetics*, **1–2**, 155–185.
- Rahmat-Samii, Y., Iversen, P. O., and Duan, D. W., 1991. GTD, PO, PTD and Gaussian beam diffraction analysis techniques applied to reflector antennas. *Applied Computational Electromagnetic Society Journal*, **6**, 6–30.
- Rahmat-Samii, Y., Huang, J., Lopez, B., Lou, M., Im, E., Durden, S., and Bahadori, K., 2005. Advanced precipitation radar antenna: array-fed offset membrane cylindrical reflector antenna. *IEEE Transactions on Antennas and Propagation*, **53**, 2503–2515.
- Rajagopalan, H., Rahmat-Samii, Y., and Imbriale, W. A., 2008. RF MEMS actuated reconfigurable reflectarray patch-slot element. *IEEE Transactions on Antennas and Propagation*, **56**, 3689–3699.
- Rudge, A. W., and Adatia, N. A., 1978. Offset-parabolic-reflector antennas: a review. *Proceedings of the IEEE*, **66**, 1592–1618.
- Rudge, A. W., Milne, K., Olver, A. D., and Knight, P. (eds.), 1982. *The Handbook of Antenna Design*. London: Peregrinus. Vol. I and II.
- Rusch, W. V. T., and Potter, P. D., 1970. *Analysis of Reflector Antennas*. New York: Academic.
- Rusch, W. V. T., Prata, A., Jr., Rahmat-Samii, Y., and Shore, R. A., 1990. Derivation and application of the equivalent paraboloid for classical offset Cassegrain and Gregorian antennas. *IEEE Transactions on Antennas and Propagation*, **38**, 1141–1149.
- Schell, A. C., 1966. The multiplate antenna. *IEEE Transactions on Antennas and Propagation*, **14**, 550–560.
- Schrank, H. E., 1985. Low sidelobe reflector antennas. *IEEE Antennas and Propagation Society Newsletter*, **27**, 5–16.
- Silver, S. (ed.), 1949. *Microwave Antenna Theory and Design*. New York: McGraw-Hill.
- Skolnik, M. I. (ed.), 1970. *Radar Handbook*. New York: McGraw-Hill. Chapter 9.
- Williams, W. F., 1965. High efficiency antenna reflector. *Microwave Journal*, **8**, 79–82.

- Wu, S. C., and Rahmat-Samii, Y., 1991. Average pattern and beam efficiency characterization of reflector antennas with random surface errors. *Journal of Electromagnetic Waves and Applications*, **5**, 1069–1087.
- Xu, S., and Rahmat-Samii, Y., 2009. A compensated spherical reflector antenna using sub-reflectorarrays. *Microwave and Optical Technology Letters*, **51**, 577–582.
- Xu, S., Rahmat-Samii, Y., and Imbriale, W. A., 2009. Subreflectorarrays for reflector surface distortion compensation. *IEEE Transactions on Antennas and Propagation*, **57**, 364–372.
- Zucker, H., 1968. Gain of antennas with random surface deviations. *Bell System Technical Journal*, **47**, 1637–1651.

Cross-references

Electromagnetic Theory and Wave Propagation
Emerging Technologies, Radar
Microwave Horn Antennas
Microwave Radiometers
Microwave Radiometers, Conventional
Radars

REMOTE SENSING AND GEOLOGIC STRUCTURE

Vernon H. Singhroy¹ and Paul Lowman²

¹Applications Development Section, Natural Resources Canada, Canada Centre for Remote Sensing, Ottawa, ON, Canada

²NASA Goddard, Code 698.0, Greenbelt, MD, USA

Introduction

Geology is sometimes distinguished from geophysics as being the study of the *outside* – the surface of the Earth and structures expressed at the surface – and geophysics being the study of the *inside* of the Earth. These light-hearted definitions are a convenient introduction to what has become the most pervasive and important direct uses of satellite images for geology and, in particular, structural geology.

The value of satellite images for topographic mapping and military reconnaissance was realized at an early stage, but there was virtually no appreciation of the value of orbital methods for geology until the mid-1960s, after hundreds of satellites had been launched. However, geologic remote sensing was given a sudden and unpredicted jump start by the US Mercury and the Gemini programs.

Project Mercury, the first American manned space effort, began in 1958, with the successful orbital flight by John Glenn in 1962. Beginning with the second flight, by Scott Carpenter, the Mercury pilots carried out terrain photography for geologic purposes. The Mercury terrain photographs generated the beginnings of wide appreciation of the geologic value of orbital photography. By the time the 10 Gemini flights were over, some eleven hundred 70 mm color photographs suitable for geology, geography, or oceanography study had been acquired. Published widely, these spectacular pictures generated worldwide interest among public and scientists alike.

Their most important result for geology was the stimulus to the US Geological Survey's Earth Resources Observation Satellite (EROS) program.

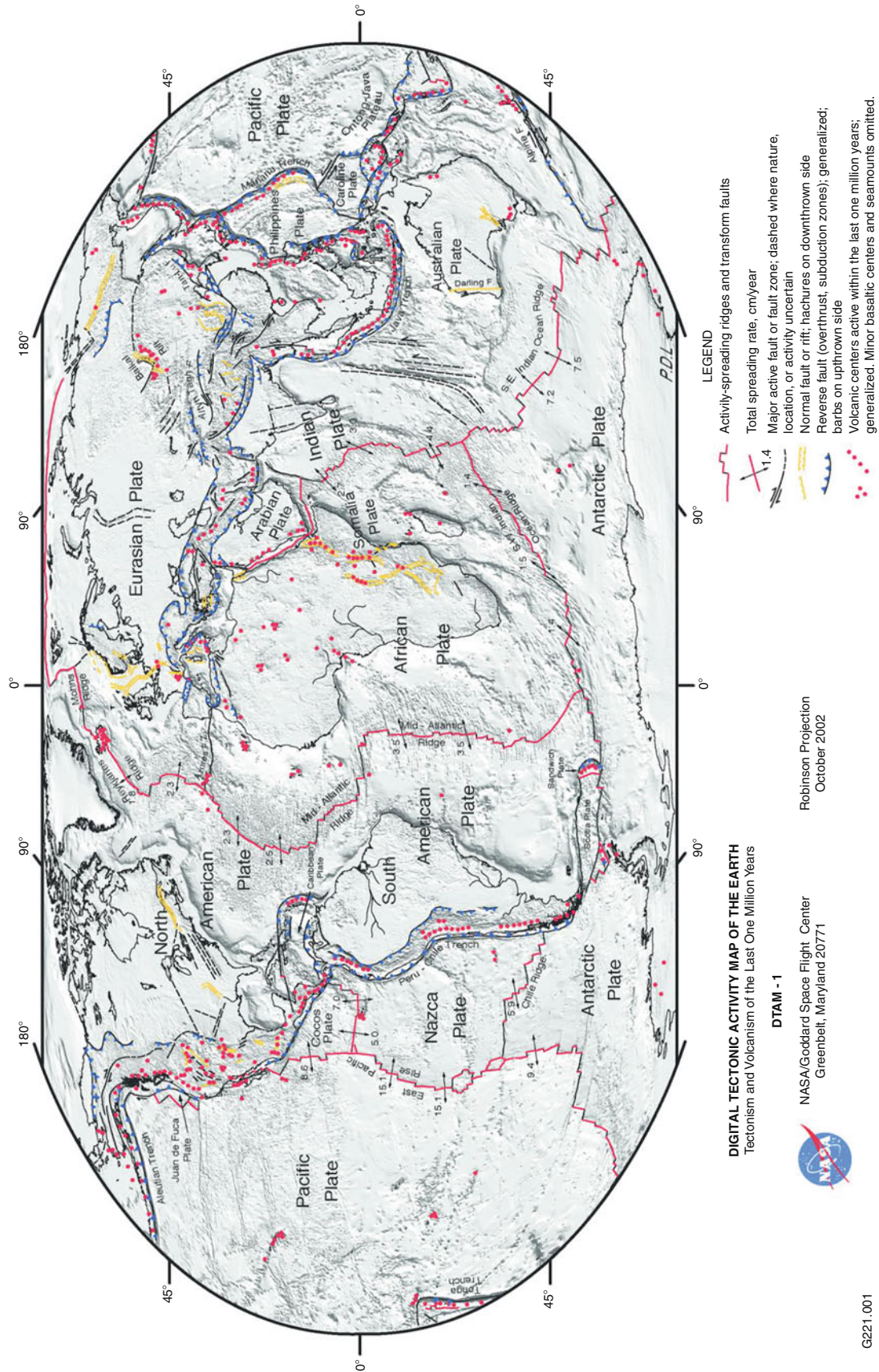
The EROS program eventually became Landsat, the first of which was launched in July 1972 and was taken over by NASA. The great success of Landsat and later missions in the series triggered similar programs in other countries, notably France, India, and Russia. In recent years, commercial programs of orbital remote sensing have been highly successful, and anyone interested can now obtain orbital images from Google Earth or other sources.

Orbital imaging radar began with the Seasat mission of the 1960s, whose results like Landsat stimulated development of successive radar-carrying satellites by the European Space Agency (ERS-1 and ERS-2). Most important was Canada's RADARSAT program, the first operational orbital radar system. RADARSAT has now produced global imagery being used for many applications, including oceanography, sea-ice observations, and geology, in particular structural geology.

"Structural geology" is equivalent to tectonics, when applied to regional or continental areas. Plate tectonics is a widely accepted theory in which the Earth's crust is considered to be a number of "plates," rigid segments of the lithosphere that are in motion relative to the interior of the Earth. Examples of satellite remote sensing applied to structural geology or tectonics are described below.

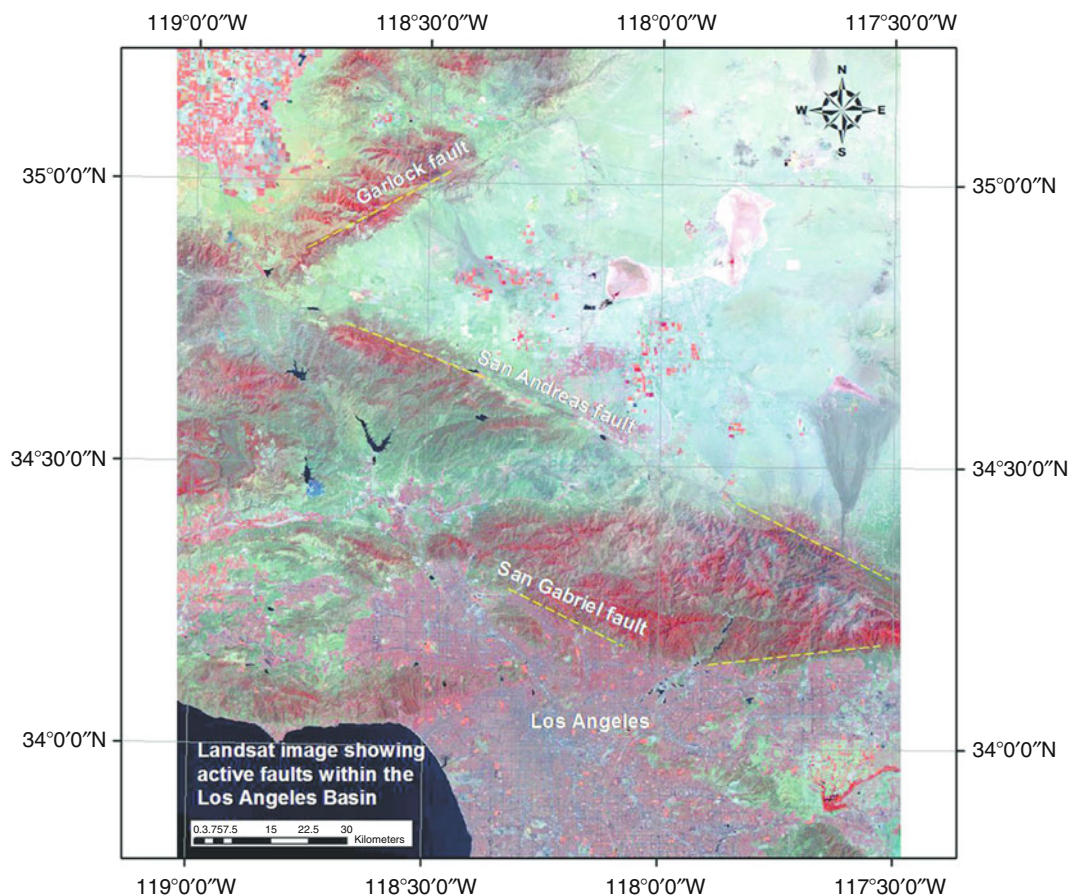
Digital tectonic activity map of the earth

The NASA Crustal Dynamics Program of the 1970s applied space geodesy techniques (satellite laser ranging and very long baseline interferometry) to the measurement of plate motions and rigidity. To plan baselines and station locations, a global map of present tectonic activity was required. The National Geographic "Physical World" map of 1975 was the base for such a map in 1977, which was used for several years by NASA for space geodesy research. Since then, the map has undergone several revisions and has appeared in 16 different textbooks and many technical papers. The most recent version, published in 2002, is a Robinson projection digital map based on sea surface altimetry for the ocean basins (Figure 1). Spreading rates are those of NUVEL-1. Terrestrial features were drawn from published literature, orbital photographs, and seismic epicenters. The one million-year figure was chosen as a time long enough to be representative but short enough that physiographic features such as fault scarps and volcanoes were still recognizable. The map has been useful in teaching geology courses at all levels. It shows recognized plates but also shows that there are many major features not on distinct plate boundaries. The map is available online as "Digital Tectonic Activity Map of the Earth," <http://denali.gsfc.nasa.gov/dtam/data/ftp/dtam.jpg> (Figure 1), accompanied by complementary seismicity and space geodesy baseline maps, and as a public domain publication. It provides a good background for several examples of satellite remote sensing in structural geology.



G221.001

Remote Sensing and Geologic Structure, Figure 1 Digital tectonic activity map of the earth.



Remote Sensing and Geologic Structure, Figure 2 Landsat image showing active faults within the Los Angeles Basin.

San Andreas Fault, California

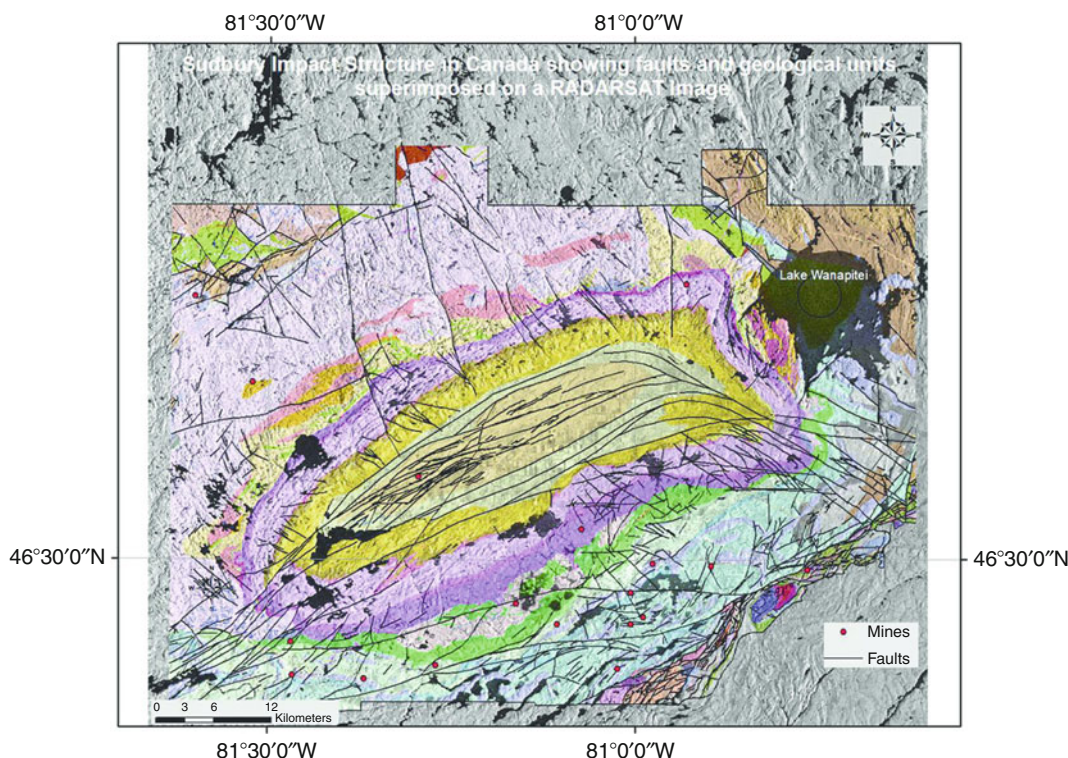
The Landsat image in Figure 2 shows the active geologic features of the densely populated Los Angeles Basin and adjacent areas. The main faults are easy to see: the San Andreas and Garlock faults outline the large triangle of the Mojave Desert. The curved south margin of the San Gabriel Mountains is another active fault. The 1971 San Fernando earthquake occurred on a fault in the San Gabriel system. All these faults are active and their location must be accurately known to construct and maintain features such as highways, pipelines, and aqueducts in the region. The fault systems and their seismic behaviors have been studied extensively. More recently ground geodetic and interferometric synthetic aperture radar satellite observations across the southern San Andreas Fault systems were used to study slip rates and the viscosity structure of the lower crust and upper mantle in Southern California.

Sudbury impact structure, Ontario, Canada

The Sudbury area in Canada is a major mineral district since the 1880s, when nickel was discovered near the city

of Sudbury. The Sudbury Basin is one of Canada's richest mining areas, with world-class mineral deposits and the world's oldest, largest, and best-exposed meteorite impact site (Mungall et al., 2004). This elliptical feature, known as the Sudbury Basin, is 300 km in diameter and approximately 2 billion years old. Over the past 100 years, over \$120 billion of nickel and copper ores were mined from more than 90 mines distributed around the rim of the Basin. The Basin has been intensely studied by geologists; it is not clear how big the structure is or how its elliptical shape was formed. Much remains to be learned about the structural evolution of the Sudbury Basin and the mineralized showings around it. The geologic structures are keys to this understanding.

Figure 3 shows some regional geologic rock units and the location of mining properties (red dots), superimposed on a RADARSAT image. The RADARSAT image helps to delineate the surface expression of mapped geologic structures published by Ames et al. (2006). RADARSAT images and image fusion techniques have been used to study the characteristic elliptical shape and the associated geologic structure fractures around the Basin (Lowman, 1994; Singhroy and Molch, 2004).



Remote Sensing, Historical Perspective, Figure 3 Sudbury impact structure in Canada showing faults and geologic units superimposed on a RADARSAT image.

The RADARSAT Standard beam mode at 20–27° incidence angles provides an excellent view of the topography and structural features of the Sudbury structure as expressed by differential erosion controlled by fractures, dykes, and lithology. The radar images have been used by geologists to study the size, shape, and evolution of the Sudbury Basin and show the result of folding from the southeast to the northwest.

Summary

Satellite images are an invaluable tool to study structural geology and regional tectonics as shown for the above examples. The synoptic and global coverage provided by satellites, the wide swath width of different orbital sensors, and the repeated coverage have produced images with different viewing angles to easily map geologic structures in vegetated and non-vegetated areas.

Bibliography

- Ames, D., Singhroy, V., Buckle, J., and Molch, K., 2006. Geology, integrated bedrock geology – RADARSAT – digital elevation data of Sudbury, ON. *Geological Survey of Canada*, Open File 4571, 1 sheet 1 CD-ROM, doi:10.4095/222241.
- Lowman, P., 1994. Radar geology of the Canadian Shield – a 10 year review. *Canadian Journal of Remote Sensing*, **20**(3), 198–209.
- Lowman, P. D., Jr., 2002. *Exploring Space, Exploring Earth*. Cambridge: Cambridge University Press, p. 362.

- Lundgren, P., Hetland, E. A., Liu, Z., and Fielding, E. J., 2009. Southern San Andreas-San Jacinto fault system slip rates estimated from earthquake cycle models constrained by GPS and interferometric synthetic aperture radar observations. *Journal of Geophysical Research*, **114**, 18.
- Mungall, J. E., Ames, D. E., and Hanley, J. J., 2004. Geochemical evidence from the Sudbury structure for crustal redistribution by large bolide impacts. *Nature*, **429**, 546–548.
- Singhroy, V., and Molch, K., 2004. Geological applications of RADARSAT-2. *Canadian Journal of Remote Sensing*, **30**(6), 893–902.

REMOTE SENSING, HISTORICAL PERSPECTIVE

Vincent V. Salomonson
Department of Geography, University of Utah, South Jordan, UT, USA

Definition

Remote sensing, historical perspective. An overview of the history of remote sensing where remote sensing is defined as the technique of obtaining information about objects through the analysis of data collected by special instruments that are not in physical contact with the objects of investigation.

Introduction

The term “remote sensing” began with the evolution of human capability to observe regions of the electromagnetic spectrum outside the range of wavelengths discernable by the human eye and traditional photography. The term “remote sensing” is attributed to Evelyn Pruitt, Office of Naval Research (see Fischer et al. in the 1975 edition of the *Manual of Remote Sensing*). Dr. Nicholas Short, formerly of NASA, provided an extended definition of remote sensing as follows:

- Remote sensing refers to instrument-based techniques used in the acquisition and measurement of spatially organized (distributed) data/information on some property(ies) (spectral, spatial, physical) of an array of target points (pixels) within the sensed scene that correspond to features, objects, and materials, doing this by applying one or more recording devices not in physical, intimate contact with the item(s) under surveillance (thus at a finite distance from the observed target, in which the spatial arrangement is preserved); techniques involve amassing knowledge pertinent to the sensed scene (target) by utilizing electromagnetic radiation, force fields, or acoustic energy through employing cameras, radiometers and scanners, lasers, radio frequency receivers, radar systems, sonar, thermal devices, seismographs, magnetometers, gravimeters, scintillometers, and other sensing instruments.

After a quick review of the beginnings of remote sensing, this overview of the history of remote sensing will be focused principally on its evolution and development for studies of the Earth-atmosphere system with emphasis on spaceborne remote-sensing systems.

Selected histories of remote sensing

Several historical perspectives can be found on the Internet and elsewhere. An excellent overview history of remote sensing by C.J. Cohen (2000) is a recommended reading. An in-depth history up through the early 1970s is provided in the *Manual of Remote Sensing*, Volume 1, pp. 27–50 (Fischer et al., 1975). Several other helpful references in the historical evolution of remote sensing include a history of NASA “Space Applications” given in Ezell (1988) or <http://history.nasa.gov/SP-4012/vol3/ch4.htm>. This resource lists key missions and also the people and documents that were important in the evolution of NASA projects designed to improve knowledge of Earth-atmosphere systems as well as providing useful information for managing the resources of the Earth. Kramer (2002) provides probably the most comprehensive listing of national and international missions and sensors along with descriptions of the background and characteristics of the missions and sensors focusing on observing the Earth and its environment. Additional insights for active and passive microwave remote sensing can be found in Njoku (1982),

Elachi (1987), and Atlas (1990). Hecht (1991) provides background history for lidar/laser remote sensing.

Historical background for Earth remote sensing

A historical background of remote sensing can be linked back to the understanding of light and its characteristics and the development and application of photography. The history can be traced back to circa 336–323 BC when Aristotle philosophized at length about the nature of light. During the intervening years, some selected events include the development of photography along with the explanation of the “camera obscura” (Latin for “dark box”) – the simplest camera by Al Hazen of Basra in 1038 AD. The principles of the camera obscura were later explained by Leonardo da Vinci in 1490. Sir Isaac Newton who was experimenting with a prism in 1666 described the dispersion of light into a spectrum of colors that could be recombined into white light. In 1839, Daguerre invented the Daguerrotype – a forerunner of pictures we know today.

Aerial photography can be traced to 1858, when Gaspar Felix Tournachon took the first aerial photograph from a captive balloon at 1,200 ft over Paris. Another noteworthy achievement was the attachment of a camera to carrier pigeons, called The Bavarian Pigeon Corps, in 1903. In 1909 at the Dresden International Photographic Exhibition, picture postcards of the fair taken by pigeons were very popular. In 1909, Wilbur Wright took an aerial photograph of Centocelli, Italy, from an airplane while trying to sell planes to the Italian government for their campaigns in Northern Africa. Photoreconnaissance began in World War I with the mounting of cameras on airplanes by the German and US armies to observe troop positions. By 1918, in the war, French aerial units were developing and printing as many as 10,000 photographs each night, during periods of intense activity. During the Meuse-Argonne offensive, 56,000 aerial prints were made and delivered to American Expeditionary Forces in 4 days. The first photograph of the actual curvature of the Earth was taken from a free balloon at an altitude of 72,000 ft by Captain Albert W. Stevens in 1936. In 1942, Kodak patented the first false color infrared-sensitive film. These events set the stage in great measures for the implementation of remote sensing from space that will be described in later paragraphs.

Some key events in development of nonphotographic remote-sensing systems should be mentioned. Notable development and discoveries include the use of the infrared portion of the electromagnetic spectrum by Sir William Frederick Herschel, a German astronomer, in 1800 (Reference: Wikipedia encyclopedia); the use of bolometers by Samuel P. Langley to make measurements of incident electromagnetic radiation or temperature in 1879; and Heinrich Hertz’s demonstration of the reflection of radio waves from solid objects in 1889 which subsequently led to the concept of radar remote sensing. In 1940, incoherent radar systems were developed in

Britain and the USA to detect and track aircraft and ships during World War II. In the 1950s, extensive research and development was carried out on infrared systems, coherent radar systems, and the synthetic aperture radar (SAR) at the University of Michigan. Major corporations such as Westinghouse, Motorola, Philco, Goodyear, Raytheon, and others developed side-looking airborne radar (SLAR) and SAR systems in the mid-1950s.

The development of laser and lidar actually was influenced significantly by the use of searchlights for remote sensing of aerosols. Hulbert (1937) pioneered the aerosol measurements using the searchlight technique and photographs of the searchlight beam to obtain aerosol observations up to a distance of 10 km. Elterman's research (1951, 1953, 1966) focused on the study of atmospheric phenomena using searchlights and made significant progress in building practical devices. The first laser – a ruby laser – was invented in 1960 by Schawlow and Townes (1958) and Maiman (1960). The first giant-pulse technique (Q-Switch) was invented by McClung and Hellwarth (1962). The first laser studies of the atmosphere were undertaken by Fiocco and Smullin (1963) and by Ligda (1963). These technologies and their applications have advanced to where lasers are now being flown in space for studies of the Earth's atmosphere, polar regions, oceans, and terrestrial ecosystems.

An overview of the evolution of Earth remote sensing from space

The beginning of Earth-system science using remote sensing from space can be traced to the launches of the early V-2 rockets (Herring and King, 2001). Following World War II, the US Naval Research Laboratory experimented with V-2 rockets and smaller ones called "sounding rockets." During a flight on March 7, 1947, what is believed to be the first space-based picture of the Earth was taken 100 miles (123 km) above the Earth over New Mexico, USA. These images and others demonstrated the potential for spaceborne cameras to help in developing a better understanding of the dynamic processes taking place in the atmosphere and on the Earth surface. The launch of Explorer 7 in 1959 included an experiment to study the Earth's radiation balance using a simple, bolometer-based sensor. Professor Verner Suomi, an eventual US Medal of Science winner, from the University of Wisconsin-Madison, was the Principal Investigator.

A major milestone for entry into the present era of sustained Earth observations from space was the launch of TIROS-1 in 1960. An excellent description of events leading to the launch of TIROS-1 is provided by Vaughan and Johnson (1994). Following the launch of the TIROS-1 satellite, the ensuing two decades (1960–1980) saw a virtual armada of satellites and accompanying sensors launched into low Earth orbit by the USA and other countries to better understand Earth-atmosphere, oceans, polar

regions, and land surface processes and trends (see Kramer, 2002). Similarly, progress was made with the launch and operation of satellites in geosynchronous orbit that provide hemispheric views of the Earth at several times during the day. These started with the NASA Advanced Technology Satellite (ATS) series beginning in 1966 and Synchronous Meteorological Satellites (SMS) starting in 1974 and evolving into the Geostationary Operational Environmental Satellite (GOES) series starting in 1975 that were operated by NOAA all using essentially a Visible and Infrared Spin Scan Radiometer (VISSR). Similar and complementary geosynchronous systems were put into place by Russia, Japan, the European Space Agency (ESA), and India, thereby providing effectively complete geosynchronous coverage of the Earth with emphasis on observing rapidly varying phenomena such as severe storms, for example, thunderstorms and hurricanes.

As time progressed, the sensors became more capable and the lifetime of the spacecraft and sensors increased (see <http://centaur.sstl.co.uk/sshp/mini/mini60s.html>). By the late 1970s and early 1980s, Earth-observing satellites and accompanying sensors often lasted for several years. As the dependability of the sensors and spacecraft improved, several nations established what were generally described as "operational" systems put in place to provide systematic, reliable, global observations of the Earth's environment with particular emphasis on meteorological variables such as cloud properties (e.g., cover and cloud-top temperature). These operational systems include now, for example, the NOAA Polar Orbiting Environmental Satellites (POES) and the Geostationary Operational Environmental Satellite (GOES) series operated by the National Oceanic and Atmospheric Administration (NOAA), the US Defense Meteorological Satellite Program (DMSP) block satellite series, the Union of Soviet Socialist Republics (USSR) Meteor series, the EUMETSAT Meteosat satellite series, the Japan Geostationary Meteorological Satellites (GMS), and India's Insat satellites (see Office of Technology Assessment, 1993).

Nimbus-7, TIROS-N, and NOAA-6. Nimbus-7 carrying the Coastal Zone Color Scanner (CZCS), the Total Ozone Mapping Spectrometer (TOMS), and the Scanning Multichannel Microwave Radiometer (SMMR) serves as an example of the progress and impact of observations from satellites. The CZCS demonstrated that biological activity in the surface waters of oceans can be monitored from space. The TOMS observations provided excellent information documenting the extent of the "ozone hole" over Antarctica. The SMMR extended microwave observations of atmosphere, oceans, and land surfaces that began with the Nimbus-5 Electrically Scanned Microwave Radiometer (ESMR). NOAA-6 was the first operational satellite (as opposed to research and development satellites such as TIROS-N) to carry the Advanced Very High Resolution Radiometers (AVHRR) that has provided since the late 1970s systematic, global observations of

a wide variety of parameters including cloud properties, [sea surface temperature](#), snow and ice cover, and vegetation dynamics and processes. Also NOAA-6 carried the TIROS Operational Vertical Sounder (TOVS) package containing the High Resolution Infrared Sounder (HIRS), the Stratospheric Sounding Unit (SSU), and the Microwave Sounding Unit (MSU) produced vertical profiles of atmospheric temperature and water vapor that could be used to complement the global radiosonde network of observations.

The Landsat satellites must be noted as having provided a major stimulus to remote sensing of the Earth's surface starting in 1972 with the launch of Landsat-1 which was then called the Earth Resources Technology Satellite (ERTS-1). Prior to the launch of the Landsat sensor systems, the photographs taken from the manned, Earth-orbiting Gemini and Apollo missions showed considerable evidence of the utility of relatively high spatial resolution images for studying land, as well as atmospheric and ocean, phenomena with some distinct advantages over airborne systems and sensors. The Landsat series of satellites demonstrated that consistent in space and time, relatively high (as compared to the spatial resolutions provided by meteorological satellite sensors) observations of the Earth's surface (i.e., less than 100 m spatial resolution) could be routinely obtained over several years for studying and documenting changes occurring on the land areas of the globe in particular (see Lauer et al., 1997; Morain, 1998; <http://erg.usgs.gov/isb/pubs/factsheets/fs02303.html>; http://landsat.usgs.gov/about_mission_history.php; http://www.centennial-offlight.gov/essay/SPACEFLIGHT/remote_sensing/SP36.htm). The Landsat series also stimulated many other relatively high-resolution systems, both by governments and the private sector around the world (presently there are several with spatial resolutions less than 1 m being provided) that have resulted in a host of applications of spaceborne remote sensing for resource monitoring, science studies of terrestrial processes and trends, and a host of commercial applications. These satellite systems include not only nationally sponsored systems by France, India, and other countries but also commercial satellite systems (see: Stoney, 1998; http://www.asprs.org/news/satellites/asprs_database_110804.pdf; http://www.eijournal.com/Markets_Opportunities.asp). At present, there are nearly 30 satellite systems in orbit according to this report and more are planned for the future by various countries.

Lastly, great progress has been achieved since the launch of TIROS-1 in the general area of satellite geodesy wherein observations of crustal movements and changes in the length-of-day of the Earth are obtained using satellite laser ranging (SLR), very long baseline interferometry (VLBI) observations, and global positioning systems (GPS) technologies. In addition highly scientifically useful observations of the Earth's gravity field and magnetic field are also obtained from space (Salomonson, 1995). Satellite geodesy is the measurement of the form and

dimensions of the Earth, the location of objects on its surface, and the figure of the Earth's gravity field by means of satellite techniques (reference: Wikipedia). Scientific papers advocating the use of satellites for geodetic purposes were published as early as 1956. Geodetic applications were outlined by the Smithsonian Astrophysical Observatory for data obtained from Project Vanguard during the 1958–1959 International Geophysical Year (IGY). With this information, the constant growth of space technology, the development of electronic distance measuring devices, and the perfection of electronic data processing equipment, satellites specifically equipped for geodetic purposes have been developed and launched, to provide both scientific and practical information for use in surveying, navigation, air/sea/land transportation, etc. Early satellite tracking involved the use of special cameras and included the tracking of Sputnik 1 and 2 in 1957 and the Vanguard satellite orbit in 1959. With the invention of the laser, the capability grew to track with greater precision through the use of laser ranging. Satellite systems with corner-cube reflectors were launched in the 1960s including the first Laser Geodynamics Satellite (LAGEOS-1) and the French Starlette satellite. The combination of satellite laser ranging (SLR) and very long baseline interferometry (VLBI) demonstrated that plate tectonic measurements could be done with accuracies in the centimeter range. The launches of Skylab in 1973 and the Geodetic and Earth-Orbiting Satellite (GEOS-3) with radar altimeters aboard yielded significant, early advances in the use of altimetry over the oceans and the mapping of the gravity field, particularly the marine geoid. Observations of the Earth's magnetic field started with the launch of Sputnik 3 and Vanguard 3 in 1958. These were followed by two Cosmos satellites in 1964 and three US Orbiting Geophysical Observatories (OGO 2, 4, and 6) between October 1967 and July 1971. Magsat launched in 1979 carried both scalar and vector magnetometers. Present space-based geodetic measurement techniques include, for example, the geodetic use of global positioning satellite systems such as the US Navstar GPS and the Russian Global Navigation Satellite System (GLONASS), laser ranging to satellites (SLR) such as the Laser Geodynamic Satellites (LAGEOS) satellites, satellite radar altimetry (mostly over oceans) using the NASA/French TOPEX/Poseidon and Jason series of satellites, orbital tracking of satellites for determining the Earth's gravitational field using the CHALLENGING Minisatellite Payload (CHAMP) operated by Germany, satellite-to-satellite tracking such as done by the NASA Gravity Recovery And Climate Experiment (GRACE), and satellite gradiometry (measuring the gravity gradient from orbit) employing the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) that is in development by the European Space Agency to be launched in 2008. In the case of the Earth's magnetic field, the Ørsted and CHAMP satellites which were launched in 1999 and 2000, respectively, have provided useful data.

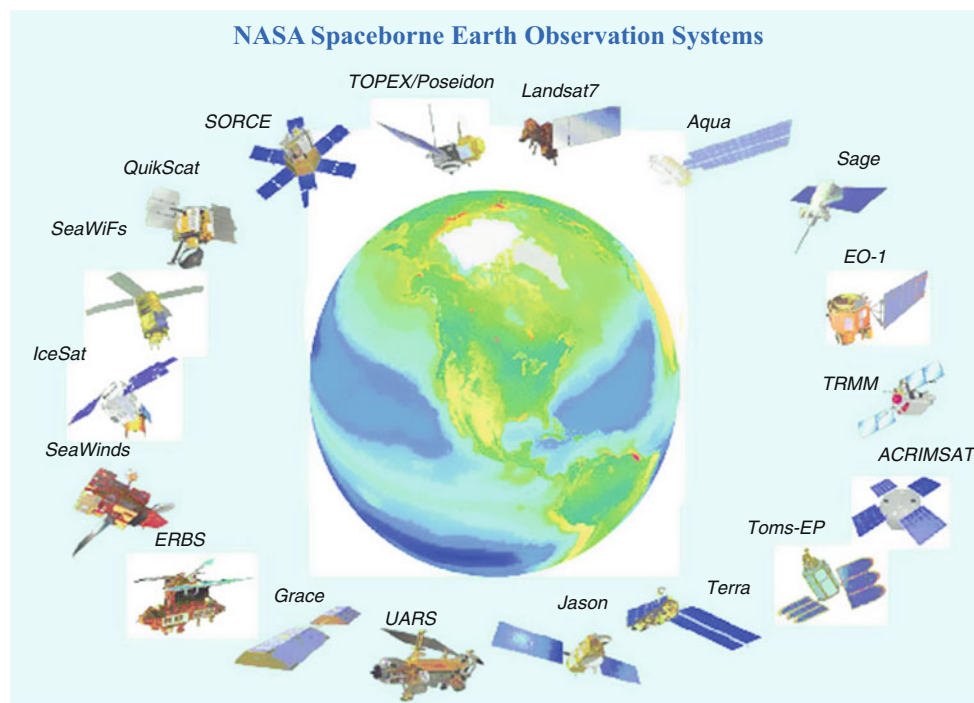
The NASA Earth-observing system (EOS), international systems, and beyond

The satellites described above and associated sensors, along with those developed and operated by other countries, produced data and information that demonstrated Earth-observing satellites can and do provide long-time series of observations that, in combination with in situ observations, result in improved insights into Earth-system processes and help decipher changes/trends (Earth System Sciences Committee, 1988). Furthermore, progress was being made by showing that these observations can be used in numerical models, for example, coupled general circulation models (GCMs), so as to improve their performance for both understanding and predicting/projecting future changes in Earth's atmosphere, oceans, land, and climate system (see National Research Council, 1991).

In addition, in the 1970s and into the 1980s, a widely held concern emerged that humankind was more than just a passenger and observer on "Spaceship Earth." There was increasing evidence that humankind was not only making greater demands on Earth's resources due to population increases and the desire for a better quality of life but also human activities could be affecting processes occurring in the land/ocean/atmosphere systems in ways that could change the composition of Earth's atmosphere, chemistry of its oceans, and composition of its land surfaces with subsequent impacts on Earth's climate and environment.

Eventually, NASA proceeded to join their Earth science efforts with the human spaceflight program by

defining very large, comprehensive, observational missions that would make use of the considerable cargo launch capabilities of the Space Shuttle and unique inclined orbit of the Space Station. These efforts were initially designated "System Z." However, these efforts evolved further to become largely independent of the NASA human space flight program and became part of NASA's "Mission to Planet Earth" in 1987 (Taubes, 1993; Pielke, 2000). The space segment of this effort was called the NASA Earth-Observing System (EOS). The early versions of the EOS envisioned a large effort composed of approximately 15 missions building and based upon the demonstrated success of previous satellite observations for studying the Earth. This set of instruments within the EOS would allow comprehensive observational efforts ranging from observing the electric and magnetic fields of the Earth to improved instruments for acquiring observations of atmosphere, oceans, and land processes and trends that would extend and improve upon data sets acquired by sensors on the Nimbus series and other operational satellite systems. In addition, new, but highly desirable instruments, such as a synthetic aperture radar (SAR) and a high spectral resolution, hyperspectral instrument (later called High Resolution Imaging Spectroradiometer [HIRIS]), could be developed and launched into space. However, over time budgetary constraints eventually reduced the original concepts to the series of missions depicted in Figure 1. The original EOS-A mission evolved to being designated "EOS-AM-1" because it had a descending



Remote Sensing, Historical Perspective, Figure 1 NASA Earth-Observing Systems flown since 1997 through 2003.

(north to south) polar orbit with an equator-crossing time near 10:30 a.m. The EOS-AM terminology was later changed to “EOS-Terra” indicating the principal emphasis of the mission was related to studies of land processes. Similarly the EOS-B mission was changed to “EOS-PM-1” indicating an equator-crossing time near 1:30 p.m. and an ascending (south to north) polar orbit. This designation was later changed to “EOS-Aqua” because the principal theme of the mission was related to the hydrological cycle, the water budget, and dynamics of the Earth-atmosphere system. The Terra mission was eventually launched in December 1999 and the Aqua mission in May of 2002. The missions depicted in Figure 1 extending from 1997 to 2003 had other missions added as time went on. The EOS Aura (focused on chemistry of atmosphere), for example, was launched in 2004. Cloudsat (cloud profiling radar observations) and CALIPSO (cloud aerosol lidar and infrared pathfinder satellite observation) missions were launched in 2006. The extent and scope of the Earth-Observing System (EOS) along with established international Earth-observing capabilities summarized above truly became a total system that is ongoing and demonstrating the power of remote sensing for gaining better understanding of Earth-system trends and processes.

Figure 1 dramatically illustrates the breadth and depth of the contributions that remote sensing is making across Earth-system science as represented by the NASA Earth-Observing Systems. There are a wide array of passive optical, multispectral observations of the land, oceans, and atmosphere features being acquired: for example, TOMS-EP, UARS, SAGE, and Aura (atmospheric chemistry); ACRIMSAT, ERBS, and SORCE (radiation balance); Landsat-7 and EO-1 (high-resolution observations land features primarily); Terra and Aqua (several optical sensors and microwave sensors/AMSU and AMSR-E observing land, ocean, and atmosphere parameters); and SeaWiFS (ocean color and land vegetation). Beyond those missions and sensors, there are active microwave/radar observations being obtained regarding precipitation and cloud water (TRMM and Cloudsat – not shown in Fig. 1); ocean surface winds (SeaWinds and QuikScat); ocean topography (TOPEX/Poseidon); lidar observations of ice sheets, clouds, and land features (ICESat and CALIPSO – not shown in Fig. 1); plus the dynamic Earth’s gravity field (GRACE). More extensive descriptions of these missions and accompanying sensors can be found in the EOS Reference Handbook at http://eospsoc.gsfc.nasa.gov/ftp_docs/2006ReferenceHandbook.pdf; Parkinson et al. (2006).

In early 2000, a new debate begun in the USA about the future of US Space Exploration beyond the completion of Space Station in 2010 and the retirement of the Space Shuttle. Based on the initial goal of a 15 year climate data record by EOS and the experience NASA gained by completing successfully the first phase of this program, a study was conducted by the US National

Academy of Sciences (NAS) to identify the second generation Earth-observing satellites that NASA and its sister agencies in the USA must build, with potential cooperation with its international partners (Space Studies Board, 2007).

There are several satellites beyond those noted above that have been put into operation and to be developed by NASA and its partners to fulfill the priorities identified by the NASA (http://eospsoc.gsfc.nasa.gov/eos_homepage/mission_profiles/index.php). For example, the Ocean Surface Topography from Space Missions (OSTM) is a continuation of the TOPEX/Poseidon and Jason missions and is to be accomplished jointly with France, NOAA, and EUMETSAT. This mission was launched in 2008. The Aquarius mission developed and launched jointly with Argentina has as the primary objective to measure the surface salinity in the oceans of the world. It was launched in 2011. The Suomi National Polar-orbiting Partnership (NPP) project was launched in 2011 and carries several improved instruments as a prototype for the next generation US National Polar Orbiting Environmental Satellite System (NPOESS). The Landsat Data Continuity Mission (LDCM) has the objective of extending the very successful Landsat series beyond Landsat-7 in cooperation with the US Geological Survey. It is to be launched in 2013.

There are also significant efforts underway globally in processing and fusing measurements from different sensors and satellites plus assimilating these data into GCMs (Le Marshall et al., 2007). For example, efforts are being made to utilize the combined power of satellites and sensors comprising what is termed the EOS “A-Train.” The “A-Train” constellation of satellite consists of GCOM-W1, Aqua, CALIPSO, Cloudsat, the French PAROSOL, and Aura. The Global Change Observation Mission-Water (GCOM-W1) “SHIZUKU” satellite was launched by the Japan Aerospace Exploration Agency (JAXA) on May 18, 2012. The Orbiting Carbon Observatory-2 satellite is scheduled to enter the A-Train in 2013. These satellites cross the equator within a few minutes of one another at around 1:30 PM local time. By combining the different sets of observations, scientists will be able to gain a better understanding of very complex Earth-system processes and parameters related to climate change.

Throughout the entire planning and execution phase of EOS program and to the present, there has been and is considerable international consultation and deliberation on potential partnerships with Europe, Japan, Russia, and other nations around the world to (1) share the costs of the EOS satellites by providing instruments, spacecraft, launching rockets, and/or contributing to the data and information management aspects of EOS program; (2) avoid duplication or gaps in critical measurements between/among the nations pursuing Earth observations from space; and (3) establish common formats, standards, and protocol for collection, archive, distribution, and sharing of the observations and information resulting from Earth-observation satellites.

Several ad hoc and formal committees/groups have been established to facilitate international dialogue and cooperation. Notable among them is the Committee on Earth Observation Satellites (CEOS) which consists primarily of the space agencies from around the world and later on its companion International Global Observing System Partnership (IGOSP) to promote the use of space-based observations by both traditional and nontraditional users of such observations. The membership of these groups/committees included primarily scientific, technical, and international officers of space agencies, international organizations, and public and private sector users of observations and information resulting from the satellites. The leadership of these groups has identified a need for high-level political support to sustain these space-based capabilities and to incorporate them into the operational networks of member organizations. To achieve this objective, the Group on Earth Observations (or GEO) was established to coordinate international efforts to build a Global Earth Observation System of Systems (GEOSS; <http://earthobservations.org/>). This emerging global coordination effort has the objectives of interconnecting a diverse and growing array of instruments and systems for monitoring and forecasting changes in the Earth's global environment. These efforts reflect that remote-sensing observations and particularly those from space have matured greatly and represent a capability that is crucial and essential to understand changes occurring on the Earth associated with natural and anthropogenic effects and to use the resulting information in early hazards warnings, managing Earth's natural resources, understanding and predicting Earth's environment, assessing the conditions of Earth's atmosphere, oceans, and continents to just name a few benefits.

Certainly remote-sensing observations now are being used for a wide variety of scientific and resource management activities, and this evolution will continue to grow and evolve into the future. It is believed that a combination of technological innovations in remote sensing and information and communication technologies together with an increasing demand for better understanding of how the Earth system and its climate is evolving to inform both policy and practical decision making in mitigating and/or managing the risks associated with changes will result in unprecedented use of these capabilities by the present generation and those who will follow for the rest of this century and beyond.

Summary

Since the beginning in understanding the power and potential of remote sensing, its use for studies of the Earth-atmosphere system has grown to the point where it is now a truly essential tool in scientific studies of the Earth and applications involved in monitoring and assessing the Earth's natural resources and its environment. The scope and extent of spaceborne systems is

worldwide including both operational and research systems operated by many nations as well as commercial systems. There has been considerable progress in the coordination among countries and organizations to achieve better, more efficient utilization of space assets for earth science and applications. Such efforts are now represented by the international efforts to build a Global Earth Observation System of Systems (GEOSS) having the objectives of interconnecting a diverse and growing array of instruments and systems for monitoring and forecasting changes in the Earth's global environment.

Bibliography

- Atlas, D., 1990. *Radar in Meteorology: Battan Memorial and 40th Anniversary Radar Meteorology Conference*. Boston: American Meteorological Society, p. 806. Hardbound. ISBN 0-933876-86-6.
- Cohen, C. J., 2000. Early history of remote sensing. In *Proceedings of 29th Applied Imagery Pattern Recognition Workshop*, October 16–18, 2000, pp. 3–9. doi: 10.1109/AIPRW.2000.953595.
- Earth System Sciences Committee, 1988. *Earth System Sciences: A Closer View*. Washington, DC: NASA.
- Elachi, C., 1987. *Introduction to the Physics and Techniques of Remote Sensing*. New York: Wiley, pp. 5–15.
- Elterman, L., 1951. The measurement of stratospheric density distribution with the searchlight technique. *Journal of Geophysical Research*, **56**, 509–520.
- Elterman, L., 1953. A series of stratospheric temperature profiles obtained with the searchlight technique. *Journal of Geophysical Research*, **58**, 519–530.
- Elterman, L., 1966. Aerosol measurements in the troposphere and stratosphere. *Applied Optics*, **5**, 1769–1776.
- Ezell, L. N., 1988. Space applications. In *NASA Historical Data Book, Vol. III: Programs and Projects, 1969–1978*, NASA SP-4012, Chap. 4.
- Fiocco, G., and Smullin, L. D., 1963. Detection of scattering layers in the upper atmosphere (60–140 km) by optical radar. *Nature*, **199**, 1275–1276.
- Fischer, W. A., Badgley, P. A., Orr, D. G., Zissis, G. J. et al., 1975. History of remote sensing. In Reeves, R. G. (Editor-in-Chief) *Manual of Remote Sensing*. Falls Church: American Society of Photogrammetry, pp. 27–50.
- Hecht, J., 1991. *Laser Pioneers*. Boston: Academic. ISBN 978-0-12-336030-4, 10: 0-12-336030-7.
- Herring, D., and King, M., 2001. *Encyclopedia of Astronomy and Astrophysics*. Bristol: IOP Publishing/Macmillan.
- Hulbert, E. O., 1937. Observations of a searchlight beam to an altitude of 28 kilometers. *Journal of Optics*, **27**, 77–382.
- Kramer, H. J., 2002. *Observation of the Earth and Its Environment: Survey of Missions and Sensors*, 4th edn. Berlin/New York: Springer, p. 1510. ISBN 3540423885.
- Lauer, D. T., Morain, S. A., and Salomonson, V. V., 1997. The Landsat program: its origins, evolution, and impacts. *Photogrammetric Engineering and Remote Sensing*, **63**(7), 831–838.
- Le Marshall, J., Uccellini, L., Einaudi, F., et al., 2007. The joint center for satellite data assimilation. *Bulletin of the American Meteorological Society*, **88**(3), 329–340.
- Ligda, M. G. H., 1963. Meteorological observations with a pulsed laser radar. In *Proceedings of the First Conference on Laser Techniques*. San Diego, CA: U.S. Navy ONR, pp. 63–72.
- Maiman, T. H., 1960. Stimulated optical radiation in ruby. *Nature*, **187**, 493.

- McClung, F. J., and Hellwarth, R. W., 1962. Giant optical pulsations from ruby laser. *Journal of Applied Physics*, **33**, 828–829.
- Morain, S. A., 1998. A brief history of remote sensing applications, with emphasis on Landsat. In Liverman, D. M. (ed.), *People and Pixels: Linking Remote Sensing and Social Science*. National Research Council, National Research (U.S.). Committee on the Human Dimensions of Global Change, published by National Academies Press, pp. 28–50.
- National Research Council, 1991. *Four-Dimensional Model Assimilation of Data: A Strategy for the Earth System*. Washington, DC: National Academy Press.
- Njoku, E. G., 1982. Passive microwave remote sensing of the earth—a review. *Proceedings of the IEEE*, **70**(7), 728–759.
- Office of Technology Assessment, U. S. Congress, 1993. *The Future of Remote Sensing from Space: Civilian Satellite Systems and Applications*, OTA-ISC-558. Washington, DC: U. S. Government Printing Office, p. 199 (see listing of non-US satellite systems, pp. 167–188).
- Parkinson, C. L., Ward, A., and King, M. D., 2006. *EOS Reference Handbook: A Guide to NASA's Earth Science Program and Earth Observing Satellite Missions*. Washington, DC: National Aeronautics and Space Administration, p. 291.
- Pielke, R. A., Jr., 2000. Policy history of the US global change research program: part I. Administrative development. *Global Environment Change*, **10**, 9–25.
- Salomonson, V. V., 1995. The contributions of spaceborne observing systems to the understanding of the solid earth and land surface processes. In Asar, G., and Dokken, D. J. (eds.), *The State of Earth Science from Space: Past Progress, Future Prospects – Proceedings of a Symposium Held, 1994, May 12*. Woodbury: AIP Press, pp. 1–18. ISBN 1-56396-492-9.
- Schawlow, A. L., and Townes, C. H., 1958. Infrared and optical masers. *Physical Review*, **112**(6), 1940–1949.
- Space Studies Board, 2007. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*, Committee on Earth Science and Applications from Space: A Community Assessment and Strategy for the Future. Washington, DC: Space Studies Board, Division on Engineering and Physical Sciences, The National Academies Press, p. 428.
- Stoney, W. E., 1998. The Pecora legacy—land observation satellites in the next century. In Stoney, W. E. (ed.), *Proceedings of the Pecora 13 Symposium, 1996, August 20–22, Sioux Falls, SD*. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, pp. 260–274.
- Taubes, G., 1993. Earth scientists look NASA's gift horse in the mouth. *Science*, **259**, 912–914.
- Vaughan, W. W., and Johnson, D. L., 1994. Meteorological satellites – the very early years prior to launch of TIROS-1. *Bulletin of the American Meteorological Society*, **75**(12), 2295–2305.

Cross-references

Commercial Remote Sensing
 Geodesy
[Global Earth Observation System of Systems \(GEOSS\)](#)
[Global Programs, Operational Systems](#)
[International Collaboration](#)
[Lidar Systems](#)
[Magnetic Field](#)
[Microwave Radiometers](#)
[Observational Platforms, Aircraft, and UAVs](#)
[Observational Systems, Satellite](#)
[Ocean Surface Topography](#)

REMOTE SENSING, PHYSICS AND TECHNIQUES

David L. Glackin
 Los Angeles, CA, USA

Synonyms

Earth observation; Earth remote sensing; Teledetection

Definition

Remote Sensing. The study of objects at a distance, without direct physical contact.

Introduction to remote sensing platforms

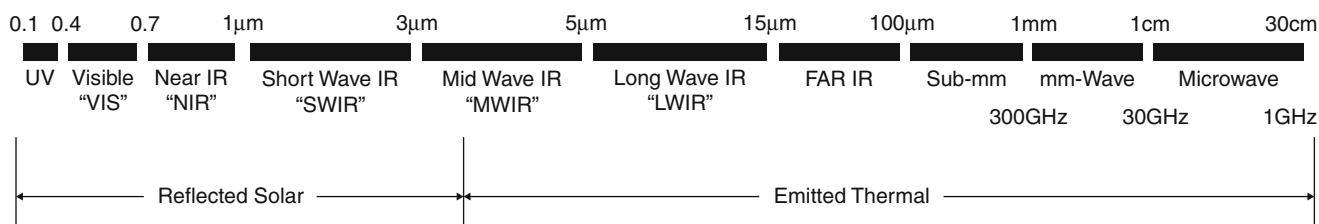
Before delving into physics and techniques, it is appropriate to briefly discuss the types of remote sensing platforms that are used to host remote sensing instruments, so that the various remote sensing techniques can be put into context. Remote sensing of the Earth can be carried out from many types of platforms, including towers, balloons, unmanned airborne vehicles (UAVs), manned aircraft, satellites, and space stations, using techniques principally in the ultraviolet (UV), visible, infrared (IR), and microwave portions of the spectrum (see [Observational Systems, Satellite](#)). Because it is impractical to instrument the entire globe with ocean buoys, land-based sensors, weather balloons, etc., the emphasis to date has been on remote sensing from aircraft and unmanned satellites.

Satellite orbits

Most Earth remote sensing satellites are placed into circular, near-polar, sun-synchronous orbits at altitudes of approximately 600–900 km above the Earth's surface. A satellite in a sun-synchronous orbit passes over a given region on the ground at approximately the same local time every day. This means that the solar illumination angle changes relatively slowly for that area from day to day, which simplifies data interpretation. Some sun-synchronous satellites are in exact repeat orbits, meaning that they fly over the same region at exactly the same local time every day. Instruments on sun-synchronous LEO satellites can cover the globe in a synoptic manner, which means they can provide coverage of large areas within a short period of time.

If an instrument onboard a satellite can gather data over a geographically wide swath (approximately 3,000 km in the direction perpendicular to the direction of satellite motion), then that instrument can cover the entire globe in 12 h, if it is designed to collect data both day and night. The visible and infrared instruments currently flying on LEO satellites typically have spatial resolutions (a measure of the size of discernible features) of approximately 1 m–1 km on the ground. Higher spatial resolution generally means less geographical coverage, owing to practical considerations of focal-plane size, onboard data storage,

David L. Glackin: deceased.



Remote Sensing, Physics and Techniques, Figure 1 Electromagnetic spectrum from the ultraviolet to the microwave.

and downlink data rates (see [Mission Operations, Science Applications/Requirements](#)).

Other remote sensing satellites, all weather satellites, occupy geosynchronous orbits. Positioned at 36,000 km above the equator, their orbital periods of 24 h make them appear to hang over one spot on Earth. With broad-area imagers on five “GEO birds,” it is possible to cover the globe up to approximately 60° latitude, with spatial resolutions as good as approximately 1 km.

The L1 libration point, a gravitationally stable point between the Earth and the sun, has been discussed as a possible location for a satellite to study the Earth as a planet for climate research. Other, less used, orbits include Medium Earth Orbit (MEO) and Molniya orbits (see [Observational Systems, Satellite](#)).

Aircraft

Despite the growth in satellite observations, aircraft are still used to gather much remote sensing data and dominate the field for some applications. Because they can fly specific paths at selected times, aircraft offer more flexibility than spacecraft, and the higher spatial resolution that is possible provides the finer detail necessary for applications such as urban planning, disaster assessment, and mapping (see [Observational Platforms, Aircraft, and UAVs](#)). Bridging the gap between conventional aircraft and orbiting spacecraft are the UAVs. These aircraft, capable of long-duration, high-altitude flight, have undergone successful tests and have the potential to provide a valuable supplement to satellites (see [Observational Platforms, Aircraft, and UAVs](#)).

The remote sensing spectrum

The portions of the electromagnetic spectrum (see [Radiation, Electromagnetic](#)) that are most useful for remote sensing can be defined as follows: The ultraviolet (UV) extends from approximately 0.1 to 0.4 μm, the visible (VIS) from 0.4 to 0.7 μm, the near infrared (NIR) from 0.7 to 1.0 μm, the short-wave infrared (SWIR) from 1 to 3 μm, the mid-wave infrared (MWIR) from 3 to 5 μm, the long-wave infrared (LWIR) from 5 to 15 μm, and the far infrared (FIR) from 15 to 100 μm. These ranges are typically defined in terms of wavelength, but other ranges can be defined in terms of frequency as well. Thus, the submillimeter-wave (sub-mm or sub-mm-wave) range

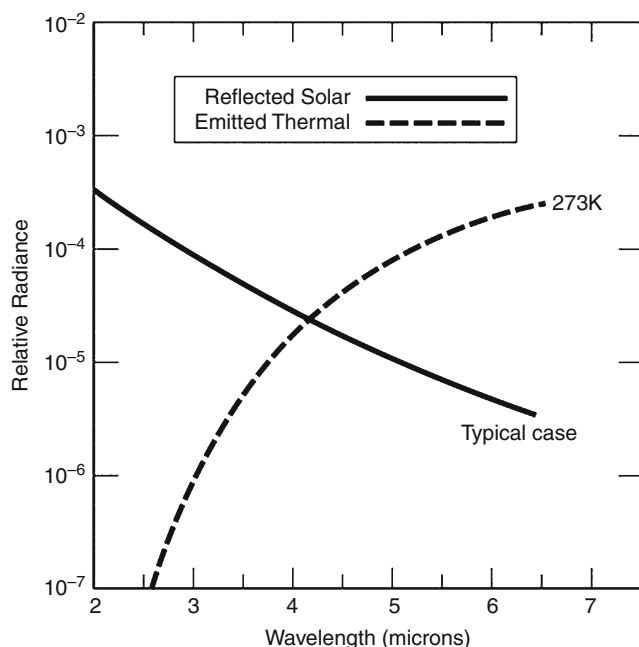
encompasses wavelengths from approximately 100 to 1,000 μm or frequencies from approximately 3 THz (THz) to 300 GHz (GHz). The millimeter-wave (mm or mm-wave) range extends from 300 to 30 GHz or 1 mm to 1 cm (cm) and the microwave region from 30 to 1 GHz or 1 to 30 cm ([Figure 1](#)).

The special case of the gamma-ray spectrum is not discussed here (see [Gamma and X-Radiation](#)).

Within these spectral regimes, there are “window bands” of low atmospheric absorption, in which imaging instruments typically operate and “absorption bands” of relatively high atmospheric absorption in which atmospheric profiling or sounding instruments operate. From the UV through the SWIR, the spectrum is dominated by reflected sunlight. In the MWIR during the daytime, the spectrum is a mix of reflected solar and emitted thermal radiation ([Figure 2](#)); the fraction of each depends on the exact circumstances of observation. From the LWIR to the FIR, the spectrum is dominated by the natural thermal emission of the environment. The sub-mm or terahertz regime (between the electro-optical and mm/microwave regimes) is in its relative infancy with regard to remote sensing techniques, showing particular potential for measuring ice in the atmosphere. The mm-wave and microwave regimes are especially useful for measuring environmental phenomena that are water related. These include atmospheric water vapor, cloud liquid water, and rainfall (see [Atmospheric General Circulation Models](#)); sea-surface temperature, wind vector, and salinity (see [Ocean, Measurements and Applications](#)); sea ice extent and concentration, snow extent, and water equivalent (see [Cryosphere, Measurements and Applications](#)); and soil moisture, land surface temperature, land freeze/thaw state, and vegetation (see [Land-Atmosphere Interactions, Evapotranspiration](#)).

The portion of the spectrum from the UV through the FIR is generally called the electro-optical (E-O) portion of the spectrum, while the portion of the spectrum from the mm-wave through microwave is generally loosely called the microwave (sometimes abbreviated as μW) portion of the spectrum.

Instruments operating in the E-O spectrum are adversely affected by clouds (see [Optical/Infrared, Scattering by Aerosols and Hydrometeors](#)) unless they are designed to be meteorological imagers or other specialized instruments. E-O imagery in the VIS-SWIR spectrum has low utility at night, although moonlight is sufficient



Remote Sensing, Physics and Techniques, Figure 2 Example of solar/thermal crossover characteristics in the mid-wave infrared. During the day, the region from 3 to 5 μm is a mix of reflected solar and emitted thermal radiation.

for high-sensitivity low-light-visible systems and hot fires are visible at night, particularly in the SWIR. E-O imagery in the LWIR-FIR is effective day or night, since it relies on natural thermal emission, not sunlight (see *Thermal Radiation Sensors (Emitted)*). Microwave instruments, on the other hand, can operate day or night and can penetrate nearly all types of weather, including most clouds. They thus offer views of areas that are frequently cloud covered, such as tropical countries, or that are in darkness for extended periods, such as the wintertime polar caps.

With the above background, an overview of remote sensing techniques and the physics upon which they rely will now be presented. The list of instruments below is not exhaustive, but it includes the major ones in use today.

Types of instruments

Remote sensing instruments fall into the general classes of passive and active electro-optical and passive and active microwave instruments. Passive instruments collect and detect natural radiation from the environment. They receive upwelling radiation from the Earth-atmosphere system, collect it, focus it, detect it, and convert it into electrical signals. These signals are then processed, stored, compressed, formatted, and transmitted to ground stations. Active instruments emit radiation and measure the returning signals. They emit a pulse of radiation that interacts with the environment and then backscatters or reflects off of it. Some of the radiation travels back to the instrument for detection and processing.

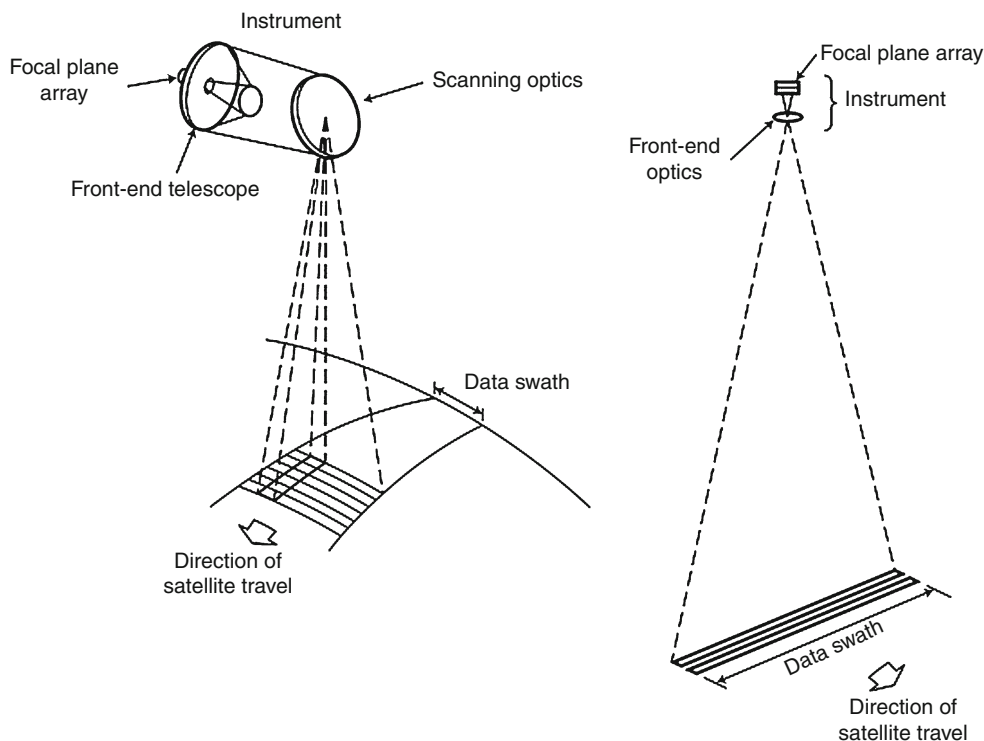
Passive electro-optical instruments include multispectral imagers, hyperspectral imagers, atmospheric profilers or sounders, spectrometers, radiometers, and polarimeters (as well as digital CCD (charge-coupled device) cameras, analog vidicon (tube-type) cameras, and film cameras). Active electro-optical instruments include backscatter lidars, differential absorption lidars, Doppler wind lidars, fluorescence lidars, and Raman lidars (“lidar” is an acronym for “light detection and ranging”). Passive microwave instruments include imaging radiometers, atmospheric sounders, spectrometers, synthetic-aperture radiometers, and submillimeter-wave radiometers (assuming for now that the sub-mm spectrum is organized under the microwave spectrum). Active microwave instruments include real-aperture radars, synthetic-aperture radars (SARs), altimeters, and scatterometers. Instruments that observe signals from GPS satellites as they propagate through the Earth’s atmosphere or scatter off the Earth’s surface, namely, GPS occultation instruments and GPS reflectometers, are organized under active microwave instruments, because the GPS signals are actively transmitted. (“Aperture,” for microwave instruments, refers to the effective size of the antenna. “Synthetic aperture” refers to the process of making the instrument function as if it has a larger or more continuous antenna than it does. For SAR, this is done by exploiting the motion of the satellite to make the antenna effectively larger. For radiometers, this is done by exploiting the properties of an array of thin antennas that only partially populate the desired aperture (“sparse aperture”), to make the aperture effectively filled in.)

The special case of using accurate measurements of the Earth’s gravitational field to infer measurements of underground water resources is not treated here. For more information, see *Water Resources*.

Passive electro-optical

There are three basic classes of passive E-O imagery: panchromatic, multispectral, and hyperspectral. Panchromatic imagery in a single broad spectral band (“black-and-white”) is useful for maximum spatial resolution, because more light is available over the wide spectral range. Multispectral imagery in several to tens of moderately wide spectral bands (“color”) adds information for discrimination, classification, and analysis of objects based on their spectral properties. Hyperspectral imagery in hundreds of narrow contiguous spectral bands further improves discrimination, classification, and analysis.

Passive electro-optical multispectral imagers observe Earth’s natural thermal radiation in the MWIR-FIR and solar radiation that has been reflected and scattered back toward space in the UV-MWIR (see *Calibration, Optical/Infrared Passive Sensors*). In the case of panchromatic or multispectral imagers looking down at the Earth from LEO, scanning optics moving in a cross-track “whiskbroom” motion, perpendicular to the direction of satellite travel, can be used to collect a swath of data.



Remote Sensing, Physics and Techniques, Figure 3 Examples of whiskbroom (left) and pushbroom (right) scan geometries. The whiskbroom scan may either be a sinusoidal scan (oscillating back and forth) or a barrel roll scan (constant direction of spin) perpendicular to the direction of flight.

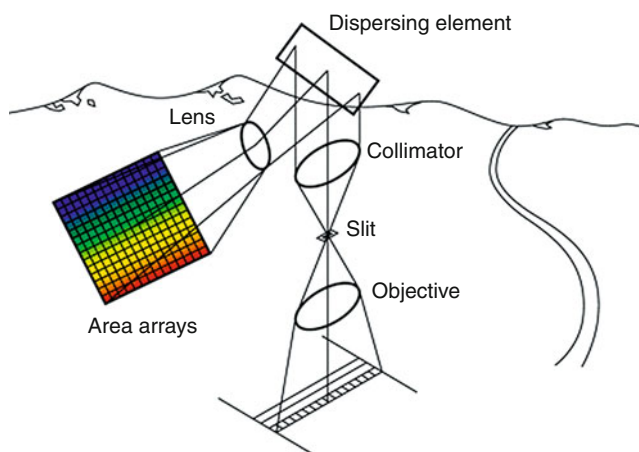
Alternatively, the motion of the spacecraft can simply carry the imager's field of view along track in a "pushbroom" fashion, parallel to the direction of satellite motion (Figure 3). The radiation captured by the front-end telescope optics is transferred through a set of back-end optics and band-pass filters to one or more focal-plane arrays, where it is converted to electrical signals by a number of detectors in the focal plane. These signals are then digitized and may be compressed to reduce down-link bandwidth requirements. Whiskbroom imagers can scan a wide swath of the planet with relatively few detectors in the focal-plane array, while pushbroom imagers can be built with no moving parts; each approach involves trade-offs (Slater, 1980). Whiskbroom scanners result in a "bow tie" effect in the data, because the projection of the focal-plane array on the Earth grows progressively larger as the instrument scans progressively farther toward the edge of the swath. The resultant data swath grows larger in both directions perpendicular to the direction of satellite motion and in outline looks like a bow tie. This effect is often removed in onboard processing.

If the multispectral imagers are calibrated to quantitatively measure the incoming radiation (as indeed most are), they are termed imaging radiometers. Such instruments detect radiation in a number of (typically less than 20) spectral bands. Multiple wavelengths are almost always required to retrieve the desired environmental

phenomena. Multispectral imagers have applications in the study of clouds, aerosols, volcanic plumes, sea-surface temperature, ocean color, vegetation, precision agriculture, forestry and deforestation, land cover and land use, urban planning, snow, ice, pollution and fire monitoring, disaster assessment, emergency response, news reporting, etc. Multispectral imagers typically fly in LEO orbits, although they are also hosted on GEO weather satellites.

Some of these applications make use of the technological advances in spatial resolution offered by the high-resolution (1 m panchromatic or less) commercial remote sensing satellites that began flying in LEO orbits in 1999. The emphasis on higher-spatial-resolution systems for mapping and other applications is sparking an increasingly close tie between remote sensing and the geographic information system (GIS) and geoprocessing technologies. GIS allows the combination of remote sensing imagery with infrastructure information such as roads, bridges, waterways, power lines, and pipelines. The ongoing increases in spatial and spectral resolution, data rate, and data volume emphasize the need for improved image compression techniques, data fusion, image archiving and browsing techniques, and advanced computer hardware (Glackin and Peltzer, 1999).

In contrast to multispectral imagers, hyperspectral imagers (HSI) typically cover 100–200 spectral bands,



Remote Sensing, Physics and Techniques,

Figure 4 Conceptual example of a hyperspectral imager, showing the implementation that uses an area array detector, in which one direction along the array is the spatial dimension and the other direction along the array is the spectral dimension. Spectra of each ground footprint along a line are collected simultaneously, and then the instrument pushbrooms along the direction of satellite motion to collect spectra of the next line on the ground. The result is a two-dimensional “image cube” of data, with two spatial axes and one spectral axis.

producing simultaneous imagery in all of them (Figure 4). Moreover, these narrow bands are usually contiguous, typically extending from the visible through SWIR regions. This makes it easier to discriminate surface types by exploiting fine details in their spectral characteristics. Exploitation of imagery from HSI systems is done with spectroscopy, whereas exploitation of imagery from multispectral systems is often done with multivariate classification. Hyperspectral imagery is used for mineral and soil-type mapping, precision agriculture, forestry, and other applications. Because the desired spatial resolution is usually on the order of 30 m, HSI instruments typically fly in LEO orbits. A few airborne hyperspectral imagers operate in the thermal (MWIR-LWIR) infrared, but this technique has yet to advance to space.

Profilers or sounders exploit the properties of a spectral band characteristic of a particular species of gas (e.g., the 15 μm band of carbon dioxide). Typically operating in the thermal infrared, they are most often used to measure the vertical profile of atmospheric temperature, moisture, ozone, and trace gases. Sounding while looking down at the Earth exploits absorption bands in the spectrum, in which the gaseous species of interest absorbs the radiation that is upwelling from the Earth. The vertical profile of the atmospheric property or species of interest can be inferred by observing in spectral regions that are narrow compared to the absorption bandwidth. When the spectral region observed is in the center of the absorption band, the instrument can see only so far down into the atmosphere, because of the high absorption. As the wavelength being observed is increasingly shifted from the absorption band

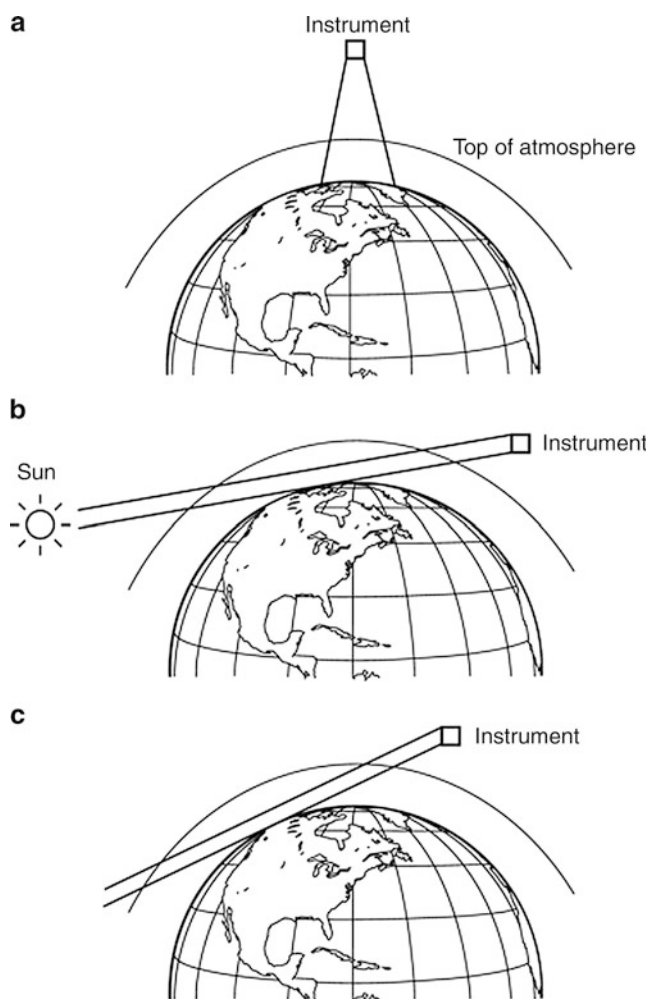
center out into the “wings” of the absorption band, absorption steadily decreases, allowing the instrument to see increasingly deeper into the atmosphere. By observing in relatively narrow portions of the spectrum, one can infer the properties of different layers of the atmosphere, because different amounts of absorption allow us to examine different depths in the atmosphere. The raw observations in the different narrow regions of the absorption band, when combined with the physics of spectroscopy and radiative transfer (see *Radiative Transfer, Theory*), and a model of the structure of the atmosphere, can yield the desired vertical profile. This process is known as “inversion”.

Sounding while looking above the edge of the Earth at the atmosphere is known as “limb sounding” (see *Limb Sounding, Atmospheric*). Limb sounding can be done by observing an atmospheric absorption band against the bright background of the solar disk (“solar occultation”). Limb sounding can also be done by observing the spectral band of a gaseous species against the black background of space, in which case the spectral band is seen in emission (i.e., a bright spectral feature against a dark background, as opposed to a dark spectral feature against a bright background) (Figure 5). Sounding may also be done with active E-O or passive microwave techniques (see below).

Spectrometers exploit the spectral “fingerprints” of environmental species and surfaces, providing much higher spectral resolution than multispectral imagers. They use a grating, prism, or more complicated method [such as a Fourier transform spectrometer (FTS) or FabryPerot spectrometer (FPS)] to spread the incoming radiation into a spectrum that can be detected and digitized. Spectrometers are typically used for measuring trace species in the atmosphere or the composition of the land surface. Various types of vegetation, soil, and minerals have unique spectral signatures.

A special class of spectrometer is the gas correlation spectroradiometer (GCS), which is similar to the pressure-modulated radiometer (PMR). The GCS is a simple, compact, and robust method for spectral detection. It uses an optical cell filled with a sample of the gas to be detected, and it compares the spectral signature of that gas with the spectral signature being observed in the Earth’s environment. Very high discrimination is possible in the presence of interference from other species. Another special class of spectrometer is the acousto-optical tunable filter (AOTF). The AOTF is an electrically tunable band-pass filter. The observed radiation from the Earth is passed through a crystal. An oscillating radio-frequency signal is applied to piezoelectric transducers (PZTs) attached to the crystal. The PZTs vibrate and generate sound waves in the crystal that turn it into an optical diffraction grating by locally perturbing the refractive index of the material. This is a compact and effective technique for spectral discrimination.

The distinction between the various classes of instruments is often blurred. For example, a sounder might use band-pass filters to observe discrete spectral bands or it



Remote Sensing, Physics and Techniques,

Figure 5 Conceptual example of (a) vertical sounding and limb (both (b) solar occultation and (c) emission) sounding.

might employ a spectrometer to observe a spectrum from which the appropriate sounding frequencies can be extracted. Similarly, a hyperspectral imager will typically use a spectrometer for spectral discrimination (in which case, it is known as an imaging spectrometer).

Nonimaging radiometers are typically used to study the Earth's energy balance. They measure radiation levels across the spectrum from the ultraviolet to the far infrared, with low spatial resolution. They can measure such quantities as the incoming solar irradiance at the top of the atmosphere and the outgoing thermal radiation caused by the sun's heating of the planet (see *Thermal Radiation Sensors (Emitted)*). These are two of the principal quantities that determine the net heating and cooling of the Earth. Such instruments typically fly in LEO, although they have application to GEO and L1 orbits.

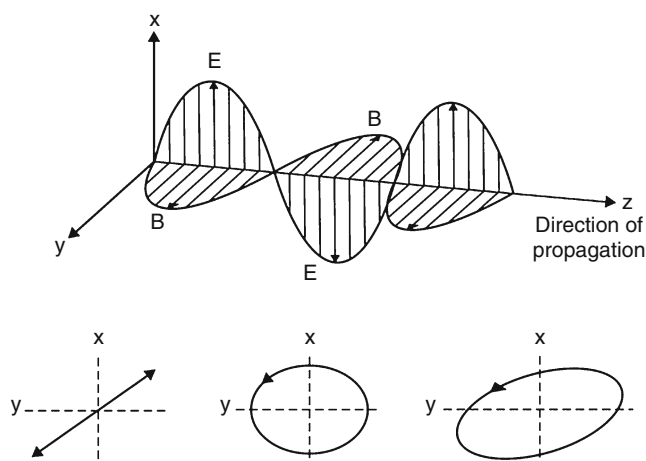
A special technique for the study of the Earth's energy balance makes use of earthshine from the Moon.

Earthshine is most easily visible to the eye when the Moon is a crescent, and it can be seen that the "dark" portion of the moon is actually faintly illuminated. This illumination is due to sunlight reflected from the Earth, off the dark portion of the Moon, and back to the Earth. Measurement of earthshine provides a measurement of the Earth's global albedo, a measure of the fraction of incident sunlight that is directly reflected back into space. Changes in the global albedo infer changes in the Earth's energy balance, which relate to climate change. This measurement is currently made from Big Bear Solar Observatory in California. This is an example in which ground-based observations looking up at the Moon have an advantage over satellite-based observations looking down at the Earth and attempting to measure the same quantity. The ground-based lunar observation measures global albedo, while the satellite-based Earth-observing instruments measure the albedo over a fraction of the globe at any given time (Goode, 1998).

Polarimeters, which can be imaging or nonimaging devices, exploit the polarization signature of the environment (see *Reflected Solar Radiation Sensors, Polarimetric*). As anyone who has experimented with polarized sunglasses knows, the sky on a clear day is polarized, particularly at a 90° angle from the sun. Sunlight reflected from metal or glass structures can also be polarized, as drivers with polarized sunglasses know. Electromagnetic radiation can be characterized as a wave, having both electric and magnetic fields, each of which can be expressed as a vector that is perpendicular to the direction of the motion of the wave (Figure 6). If the direction of the electric vector varies randomly with time, the radiation is said to be unpolarized. If the vector has a preferred orientation as a function of time, the radiation is polarized. Radiation from Earth-atmosphere system can be unpolarized, linearly polarized, circularly polarized, or elliptically polarized, depending on the physics of reflection and scattering. Polarization is fully characterized by the Stokes parameters (see *Reflected Solar Radiation Sensors, Polarimetric*). The resulting information can be used to study phenomena such as cloud-droplet size distribution and optical thickness, aerosol properties, vegetation, and other land surface properties.

Stereo imagers are a special case of E-O instruments and fly in LEO orbits. Stereo imagers that afford both fore and aft looks as the satellite passes over the Earth can provide "3-D" data useful for topographic mapping. A nadir (straight down) look is often added for accuracy in mountainous terrain (see *Land Surface Topography*). Multi-angle imagers are also a special case, often having many more view angles than three. When implemented as a polarimeter, multi-angle instruments are particularly useful for measuring the properties of clouds and aerosols (see *Reflected Solar Radiation Sensors, Multiangle Imaging*).

UV instruments are also something of a special case (see *Ultraviolet Sensors*). Their observations are limited to the Earth's upper atmosphere and near-Earth space



Remote Sensing, Physics and Techniques, Figure 6 Diagram showing the electric (E) and magnetic (H) vector in an electromagnetic wave (top). For linear polarization (left), the direction of the E -vector is fixed. For circular polarization (center), the direction of the E -vector, when viewed along the direction of propagation of the electromagnetic wave, continually moves in a circle (right- or left-hand). For the general case of elliptical polarization (right), the E -vector moves in an elliptical manner.

environment (Huffman, 1992). UV instruments are used for ozone monitoring (see *Stratospheric Ozone*) and for imaging of the Earth's aurora, airglow, polar mesospheric clouds ("noctilucent" clouds), and ionosphere.

Calibration of passive electro-optical instruments is a very important issue, if their raw observations are to be turned into scientifically and operationally useful data (National Research Council, 2007). The four types of calibration are radiometric, spectral, geometric, and polarimetric. Radiometric calibration is important both in absolute terms (conversion of raw counts to absolute radiance or irradiance units, e.g., for temperature measurement) and in relative terms (for combining observations in different spectral bands). Spectral calibration, the determination of the exact positions in the spectrum at which the data were taken, must be performed for spectrometers and imaging spectrometers. Geometric calibration is performed to determine the position on the ground at which data is taken and to remove instrumentally induced distortions. Polarimetric calibration, the determination of the type and percent polarization that is observed, must be performed for polarimeters (see *Calibration, Optical/Infrared Passive Sensors*).

Active electro-optical

A lidar sends a laser beam into Earth's environment and measures what is returned (Figure 7) via reflection and scattering. This typically requires a large receiving telescope to capture the returning photons. The returning signal can be measured either by direct detection or by heterodyne detection (NASA ESTO, 2006). With direct detection, the receiving telescope acts as a simple light

bucket to collect photons, which means that the phase information in the electromagnetic wave is lost. With heterodyne ("coherent") detection, the returning photons are combined with the signal from a laser onboard the satellite (a "local oscillator"), whose frequency is close to that of the lidar's frequency. The act of combining these two signals generates an "intermediate frequency" that is much lower in frequency than either the lidar or the local oscillator (like a heterodyne circuit in a radio), making it easier to detect while maintaining the frequency and phase information.

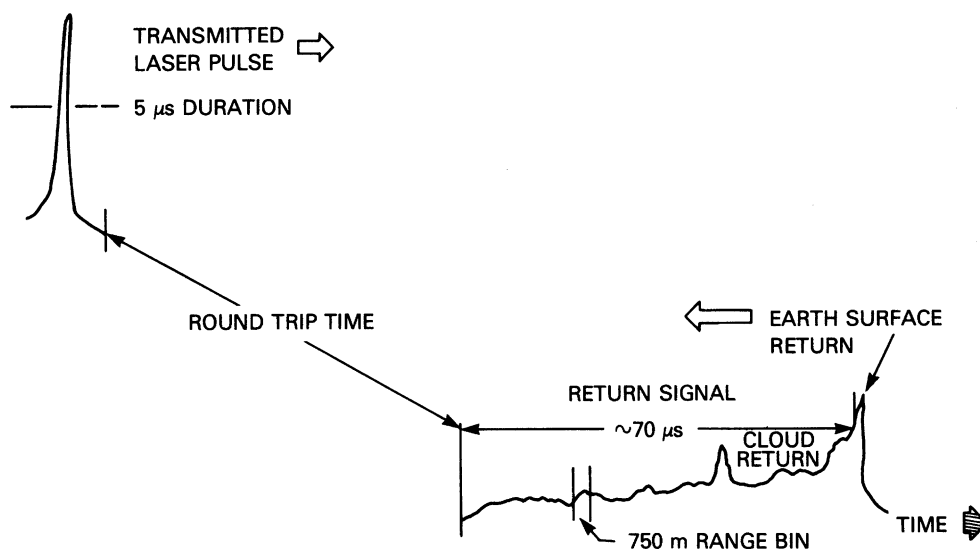
Only a handful of lidars have flown in space, owing to limitations involving high power, high cost, and the availability of robust laser sources. With a few notable exceptions, lidar remote sensing is typically carried out with an aircraft.

Lidars can potentially generate high-resolution vertical profiles of atmospheric temperature and moisture because the returns can be sliced up or "range gated" in time (and thus space) if they are strong enough. Lidar also has potential for profiling winds, determining cloud physics, measuring trace-species concentration, etc.

Backscatter lidar is the simplest in concept: A laser beam scatters off of aerosols, clouds, dust, and plumes in the atmosphere. The data can be used to generate vertical profiles of these phenomena, except where the beam is absorbed by clouds. A related device is the laser altimeter, which records the backscatter from the Earth's surface to measure features such as ice topography and the vegetative canopy (e.g., the tops of trees for biomass studies).

Differential absorption lidar (DIAL) transmits at two wavelengths, one near the center of a spectral absorption band of interest, the other just outside of it. The difference in the returned signal can be used to derive species concentration, temperature, moisture, or other phenomena, depending on the spectral band selected (akin to passive E-O sounding). The differential technique requires no absolute calibration, so it is relatively easy to achieve high accuracy (e.g., parts-per-million to parts-per-billion for species concentration).

Doppler lidar measures the Doppler shift of aerosols or molecules that are carried along with the wind. Thus, wind speed and direction can be determined if two separate views of each atmospheric parcel are acquired to measure velocity in the horizontal plane. In concept, this can be done with a conically scanning lidar and a large receiving telescope. The available aerosol backscatter is too low to measure the complete wind profile as desired (from the surface to 20 km in altitude), but molecular scattering can be used to cover the aerosol-sparse regions. Strong competition has existed between two schools of thought that propose using direct or heterodyne detection. Although wind lidar has been studied since 1978, according to current plans, approximately 30 years would have passed before the first Doppler lidar is launched into space (Kramer, 2002). Of all of the environmental parameters that can only be crudely measured today, improved wind measurements would have the



Remote Sensing, Physics and Techniques, Figure 7 Example of a lidar pulse that is emitted from the instrument and is backscattered from the atmosphere, clouds, and the Earth's surface.

greatest impact on the accuracy of numerical weather prediction (see [Tropospheric Winds](#)).

Fluorescence lidar transmits a spectral frequency that is absorbed by the species of interest and then reradiated at a different frequency, which is then detected on orbit by a radiometer. A related technology, Raman lidar, exploits the Raman scattering from molecules in the air, a process in which energy is typically lost and the scattered light is reduced in frequency.

Calibration of lidars is a very important issue, if their raw observations are to be turned into scientifically and operationally useful data.

Passive microwave

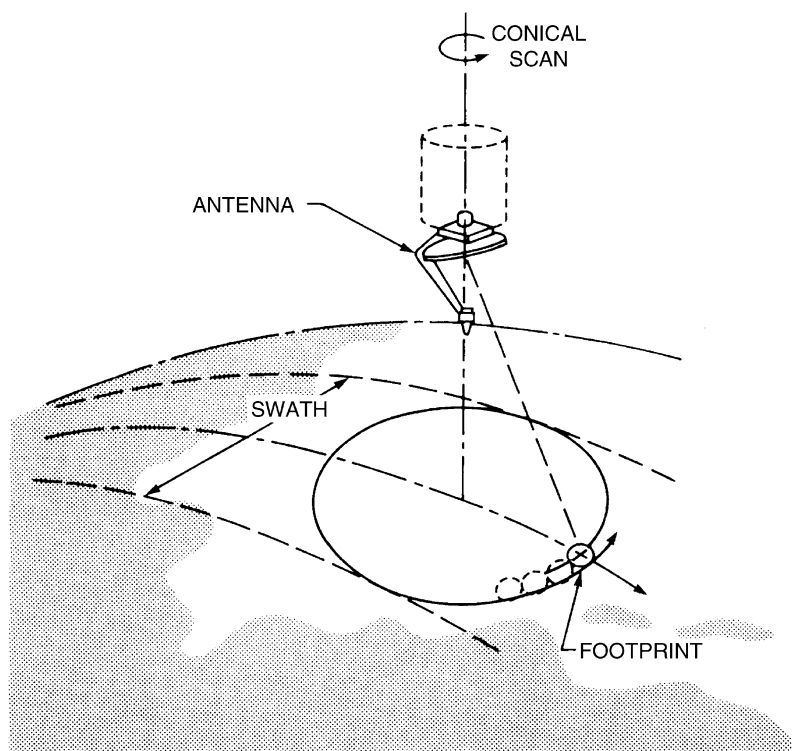
Passive microwave imaging radiometers (usually called microwave imagers; see [Microwave Radiometers, Conventional](#)) collect the Earth's natural radiation (see [Microwave Radiometers](#)) with an antenna and typically focus it onto one or more feed horns that are sensitive to particular frequencies and polarizations. From there, it is detected as an electrical signal, amplified, digitized, and recorded for the various frequencies and polarizations (linear or circular; see [Microwave Radiometers, Polarimeters](#)). The antenna usually rotates such that it scans the Earth in a conical fashion ([Figure 8](#)), and these instruments have all flown in LEO. The amount of radiation measured at different frequencies and polarizations can be analyzed to produce environmental parameters such as soil moisture content, precipitation, sea-surface wind speed, sea-surface temperature, ocean salinity, snow cover and water content, sea ice cover, atmospheric water content, and cloud water content (NASA ESTO, 2004). Unlike visible imagers, microwave imagers can operate day or night through most types of weather, including most clouds,

affording views of portions of the globe that are usually cloud covered, including the polar ice caps and many tropical countries (Janssen, 1993). Microwave instruments are ideal for imaging sea ice during the polar winter when the polar region is not illuminated by the sun and they can see through the persistent cloud cover that often plagues these regions (see Ulaby et al., 1981).

A special class of microwave radiometer is a passive polarimetric radiometer that measures not only sea-surface wind speed but also sea-surface wind direction. By measuring various combinations of linear and circular polarization from the sea surface at different frequencies (e.g., 10, 18, and 36 GHz), a signal can be derived that varies with the conically scanning look angle as a function of wind direction. This provides a potential alternative to scatterometers (see section "[Active Microwave](#)" below).

Microwave profilers or sounders, like electro-optical sounders, operate in several frequencies around a spectral band characteristic of a target gas and may be either vertical sounders or limb sounders. They are often used to measure the vertical profiles of temperature and moisture in the atmosphere. The oxygen band near the frequency 60 GHz, which becomes more or less opaque as a function of atmospheric temperature, is usually used for temperature sounding, while the water vapor band at 183 GHz is typically used for moisture sounding. The advantage of microwave over electro-optical sounding is that it can be done through most types of weather and cloud cover.

Passive microwave imagers and sounders generally operate at frequencies ranging from 6 to 183 GHz. Higher frequencies have recently been used in submillimeter-wave radiometers for measuring cloud ice content. Lower frequencies, around 1 GHz, can be used to measure soil moisture and ocean salinity; however, such low frequencies are not always practical. For a given antenna size,



Remote Sensing, Physics and Techniques, Figure 8 A conventional conically scanning microwave imager, which maps out a data swath as a result of the motion of the satellite.

spatial resolution decreases as the frequency decreases. Most microwave imagers are limited to a lower frequency of about 6 GHz because a large antenna would be required at 1 GHz to achieve acceptable resolution.

This difficulty can be overcome through a technique known as aperture synthesis (see *Microwave Radiometers, Interferometers*). In this concept, which has long been used in radio astronomy, the operation of a large solid dish antenna is simulated by using only a sparsely populated aperture or “thinned array” antenna. In such an antenna, only part of the aperture physically exists and the remainder is synthesized by correlating the individual antenna elements (see *Microwave Radiometers, Correlation*). This technique has been proven in aircraft flight demonstrations, but has not yet been flown in space. This type of instrument is often referred to as a STAR (Synthetic Thinned Array Radiometer – [Figure 9](#)).

A significant problem in the field of microwave radiometry is the existence of man-made sources of radio-frequency interference (RFI) on the Earth, especially in advanced and highly populated regions. Research on various techniques for addressing this problem is ongoing.

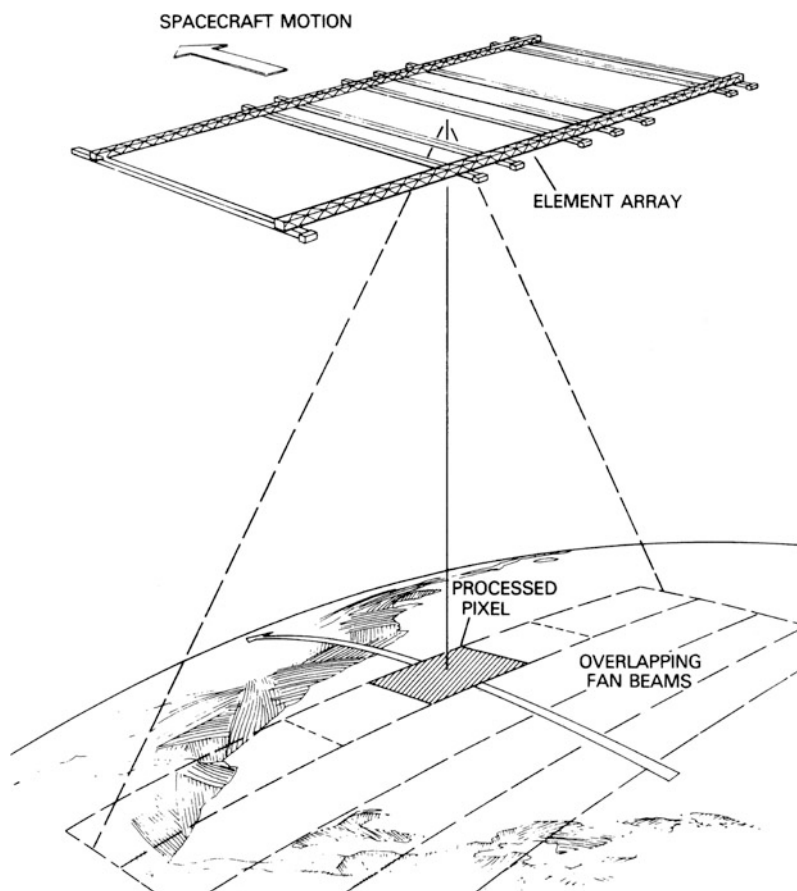
Calibration of microwave radiometers is a very important issue, if their raw observations are to be turned into scientifically and operationally useful data (Skou, 1989). Radiometric calibration is important both in absolute terms for conversion of observed radiation to temperature (“brightness temperature”) of the Earth’s environment and

in relative terms (one frequency relative to another). It is usually done in spaceborne instruments using a regulated “warm load” and a view to cold deep space, each of which is viewed periodically as the antenna and feedhorn assembly rotates past the warm load and the space view (see *Calibration, Microwave Radiometers*).

Active microwave

Active microwave instruments can be broadly divided into real-aperture and synthetic-aperture radars and operate from LEO. They all transmit microwaves toward Earth and measure what is reflected and scattered back. “Real aperture” typically means that the instrument has a solid, dish-shaped antenna that defines its performance properties, such as spatial resolution on the ground. “Synthetic aperture” usually means that the performance of a hypothetically very large real aperture is achieved by letting the satellite’s motion along its orbit sweep the field of view of the radar along with it for a period of time while data is collected, such that the length of the effective aperture equals the amount of space swept out by the satellite during that period of time (Ulaby et al., 1982; Cantafio, 1989; NASA ESTO, 2004).

Real-aperture radars can be further categorized as atmospheric radars, altimeters, and scatterometers. Atmospheric radars are useful for studying precipitation and the three-dimensional structure of clouds. The use of more



Remote Sensing, Physics and Techniques, Figure 9 Conceptual example of a Synthetic Thinned Array Radiometer (STAR) passive microwave interferometer, in which the full aperture is synthesized by correlating the signals from a small number of stick antennas.

than one frequency is beneficial for separating the effects of cloud and rain attenuation from those of backscatter. The first spaceborne atmospheric radar to fly was the precipitation radar on the US/Japan Tropical Rainfall Measuring Mission. The second was on NASA's CloudSat mission, which performed the first 3-D profiling of clouds. This mission is especially important because clouds and aerosols are the primary unknowns in the global climate-change equation.

Altimeters measure surface topography, and radar altimeters are typically used to measure the surface topography of the ocean (which is not as uniform as one might think). They operate using time-of-flight measurements and typically use two or more frequencies to compensate for ionospheric and atmospheric delays (see [Radar; Altimeters](#)). Altimeters have been flying since the days of Skylab in 1973. Aperture synthesis and interferometric techniques can also be employed in altimeters, depending on the application.

Scatterometers are a form of instrument that uses radar backscattering from the Earth's surface (see [Radar; Scatterometers](#)). The most prevalent application is for the measurement of sea-surface wind speed and direction.

This type of instrument first flew on Seasat in 1978. A special class of scatterometer called delta-k radar can measure ocean surface currents and the ocean wave spectrum using two or more closely spaced frequencies.

Synthetic-aperture radars (SARs) also flew for the first time on Seasat. These radars sometimes transmit in one polarization (horizontal or vertical) and receive in one or the other (Campbell, 2007). A fully polarimetric synthetic-aperture radar employs all four possible send/receive combinations (see [Radar; Synthetic Aperture](#)). Synthetic-aperture radars are powerful and flexible instruments that have a wide range of applications, such as monitoring sea ice, oil spills, soil moisture, snow, vegetation, and forest cover.

Some active microwave instruments (predominantly SARs) are interferometric (InSAR), meaning that they exploit the signals that are seen from two somewhat different locations but are sufficiently close in time. This is a powerful means of elevation/topographic measurement. Interferometry can be done by using two antennas separated by a rigid boom, or by using a single antenna on a moving spacecraft that acquires data at two different times or by using similar antennas on two separate

spacecraft. InSAR can be used to monitor surface motion due to earthquakes and volcanoes and to create Digital Elevation Models that represent the three-dimensional topography of the Earth's surface, information that has a wealth of applications.

GPS occultation instruments exploit a technique that was first used for remote sensing of planetary atmospheres. They observe the GPS satellites as they rise and set over the Earth's limb (NASA ESTO, 2004). They detect the signals emitted from the GPS satellites and track the phase of those signals as a function of altitude above the limb (see *GPS, Occultation Systems*). Those phase changes can be converted to atmospheric density changes, which in turn can be used to compute the temperature profile in the lower atmosphere, as well as the electron density profile in the ionosphere (the moisture profile in the lower atmosphere can also be derived, but with more difficulty). Temperature profiles derived using this technique are more accurate and structured than those using weather balloons. The other type of GPS-based instrument, the GPS reflectometer, measures the difference between direct and ocean-reflected GPS signals, to measure ocean surface height (topography). This is a newer technique that has flown on three satellites through 2008.

Calibration of active microwave instruments is a very important issue, if their raw observations are to be turned into scientifically and operationally useful data (see *Calibration, Scatterometers* and *Calibration, Synthetic Aperture Radars*).

Conclusion

Remote sensing is a field that involves a wide variety of instruments that exploit the physics of the Earth's environment (atmosphere, oceans, land, solid Earth, ice cover, and near-Earth space environment) to measure a large number of environmental phenomena. The field has blossomed tremendously since the launches of Landsat-1 in 1972, Skylab in 1973, and Seasat and Nimbus-7 in 1978. New and more sophisticated techniques are being applied, resulting in the ability to measure a wider range of environmental phenomena and an increase in the accuracy with which those measurements are made. Technologies for access to and manipulation of the data have increased dramatically. The field has benefited from international proliferation, as the number of countries owning remote sensing satellites increased by a factor of 6 from 1980 to 2008. As the field evolves, there will be a continued need for trained scientists and engineers with a broad and interdisciplinary perspective to work in the remote sensing field.

Bibliography

- Campbell, J. B., 2007. *Introduction to Remote Sensing*. New York: The Guilford Press.
- Cantafio, L. J. (ed.), 1989. *Space-Based Radar Handbook*. Norwood: Artech House.
- Glackin, D. L., and Peltzer, G. R., 1999. *Civil, Commercial, and International Remote Sensing Systems and Geoprocessing*. El Segundo, CA: The Aerospace Press and AIAA.

- Goode, P., 1998. *Earthshine measurements of global atmospheric properties*. http://www.bbso.njit.edu/Research/EarthShine/espaper/earthshine_proposal.html
- Huffman, R. E., 1992. *Atmospheric Ultraviolet Remote Sensing*. San Diego, CA: Academic.
- Janssen, M. A., 1993. *Atmospheric Remote Sensing by Microwave Radiometry*. New York: Wiley.
- Kramer, H. J., 2002. *Observation of the Earth and its Environment: Survey of Missions and Sensors*, 4th edn. Berlin: Springer.
- NASA ESTO, 2004. *Active and passive microwave technologies working group report*. http://esto.gsfc.nasa.gov/adv_planning.html
- NASA ESTO, 2006. *Working group report: lidar technologies*. http://esto.gsfc.nasa.gov/adv_planning.html
- National Research Council, 2007. *Earth Science and Applications from Space*. Washington, DC: The National Academies.
- Skou, N., 1989. *Microwave Radiometer Systems: Design and Analysis*. Norwood: Artech House.
- Slater, P. N., 1980. *Remote Sensing: Optics and Optical Systems*. Reading, MA: Addison-Wesley.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., (1981) *Microwave Remote Sensing: Active and Passive, Vol. I – Microwave Remote Sensing Fundamentals and Radiometry*, Addison-Wesley, Advanced Book Program, Reading, Massachusetts, p. 456.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., (1982) *Microwave Remote Sensing: Active and Passive, Vol. II – Radar Remote Sensing and Surface Scattering and Emission Theory*, Addison-Wesley, Advanced Book Program, Reading, Massachusetts, p. 609.

Cross-references

- Calibration, Microwave Radiometers
- Calibration, Optical/Infrared Passive Sensors
- Calibration, Scatterometers
- Calibration, Synthetic Aperture Radars
- Cryosphere, Measurements and Applications
- Gamma and X-Radiation
- Geophysical Retrieval, Inverse Problems in Remote Sensing
- GPS, Occultation Systems
- Land Surface Topography
- Lidar Systems
- Limb Sounding, Atmospheric
- Microwave Horn Antennas
- Microwave Radiometers, Conventional
- Microwave Radiometers, Correlation
- Microwave Radiometers
- Observational Platforms, Aircraft, and UAVs
- Observational Systems, Satellite
- Ocean, Measurements and Applications
- Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties
- Optical/Infrared, Radiative Transfer
- Optical/Infrared, Scattering by Aerosols and Hydrometeors
- Radar, Altimeters
- Radar, Scatterometers
- Radar, Synthetic Aperture
- Radiation, Electromagnetic
- Radiative Transfer, Theory
- Radio-Frequency Interference (RFI) in Passive Microwave Sensing
- Reflected Solar Radiation Sensors, Multiangle Imaging
- Reflected Solar Radiation Sensors, Polarimetric
- Thermal Radiation Sensors (Emitted)
- Tropospheric Winds
- Ultraviolet Sensors
- Water Resources

RESOURCE EXPLORATION

Fred A. Kruse¹ and Sandra L. Perry²

¹Physics Department and Remote Sensing Center, Naval Postgraduate School, Monterey, CA, USA

²Perry Remote Sensing, LLC, Denver, CO, USA

Synonyms

Mineral exploration Nonrenewable resource exploration; Oil and gas exploration

Definition

Use of remote sensing technology to directly or indirectly explore for occurrences of specific nonrenewable resources.

Remote sensing resource exploration

Introduction

Nonrenewable resources are a key part of daily life. Materials as varied as salt, silicon, clays, and diamonds are used in many products. Metals including lead, silver, copper, molybdenum, gold, and many others are used daily for both basic needs and in modern conveniences. Oil and gas provide power for automobiles and other transportation, lighting and heating, and manufacturing.

Classical resource exploration typically utilizes field geologic mapping of the physical characteristics of rocks and soils such as outcrop exposure, mineralogy, weathering characteristics, and geochemical and/or geophysical signatures to determine the nature and distribution of geologic units and associated resources. Structural/geomorphic interpretation, including estimation of rock orientation, mapping key unit contacts, identifying folds and faults, and interpreting drainage patterns and linear/circular features, also provides clues to help locate specific resources. Unfortunately, many of the factors important for making distinctions for resource discovery and mapping are often difficult to map in the field. Several geologists mapping the same area often produce very different geologic maps. In addition, many areas of the world have no geologic maps or incomplete geologic maps. Remote sensing technology provides a unique means of quickly mapping Earth-surface properties that can potentially improve and enhance the resource exploration process.

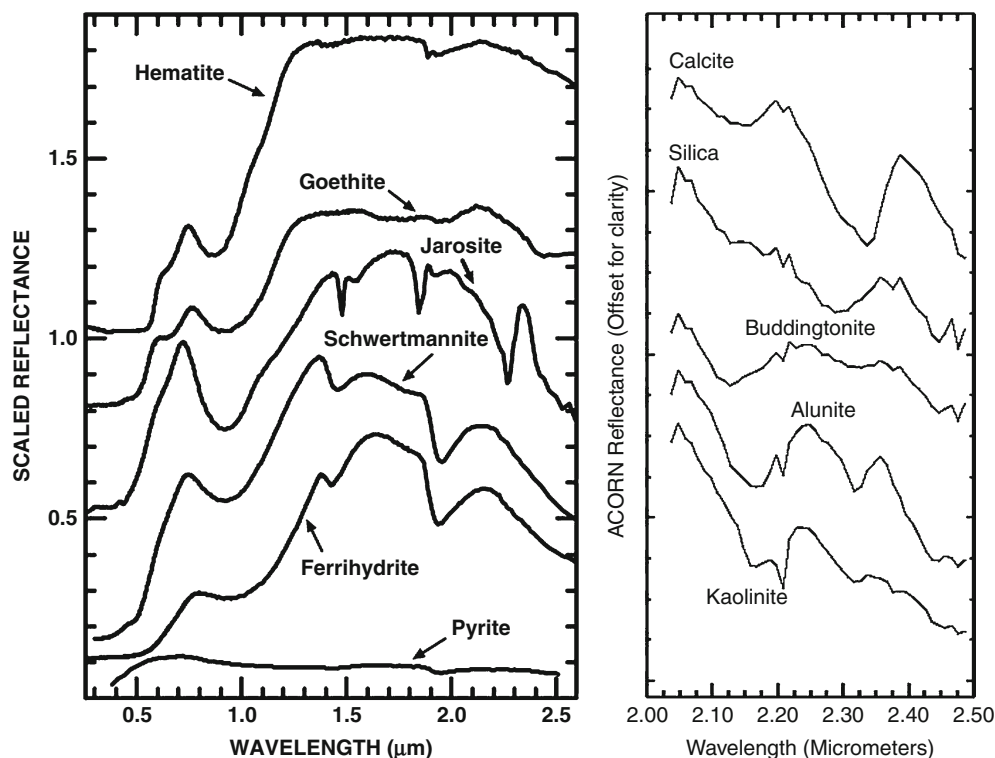
Direct versus indirect methods

Ideally remote sensing technology would directly detect the resources of interest. While this is possible for some resources (e.g., spectral mapping of a specific materials like carbonate rocks or sources of silica), most resources do not have direct indicators and use of indirect remote sensing indicators is required. Some of these indirect indicators include lithology, alteration minerals, vegetation distribution and vegetation-stress mapping, and mapping geologic structures.

In mineral exploration, direct spectral detection of ore minerals such as pyrite using remote sensing data would be an ideal method of locating, mapping, and monitoring metallic mineral deposits. Pyrite and other ore minerals cannot, however, generally be mapped using VNIR/SWIR spectral techniques (Swayze et al., 2000) (Figure 1, left). Fortunately, secondary Fe-bearing minerals resulting from the alteration or weathering of pyrite have distinctive spectral signatures in the VNIR (Taranik and Kruse, 1989; Clark et al., 1993a; Farrand, 1997; Rowan et al., 1996; Rowan, 1998) (Figure 1, left). Spectral data can be used to directly map this indirect evidence – allowing identification of even individual iron mineral species (Swayze et al., 2000). In addition, given the proper instrument (spectral and spatial resolution) and signal-to-noise (SNR) performance, spectroscopy can be used to unambiguously identify, quantify, and map other ubiquitous alteration and weathering minerals often associated with ore (minerals such as kaolinite, alunite, sericite (muscovite), montmorillonite) based on their molecular absorptions in the Short-Wave-Infrared (SWIR) (Goetz et al., 1985; Kruse, 1996; Swayze et al., 2000; Kruse et al., 2003, 2006; Mars and Rowan, 2006) (Figure 1, right). LWIR emissivity spectra also provide the ability to discriminate between important rock-forming minerals based on fundamental molecular absorptions (Lyon, 1964; Salisbury et al., 1988, 1992; Salisbury and D'Aria, 1992) (Figure 2).

Oil and gas accumulations occur primarily in sedimentary rocks. Requirements for formation of this resource include source rocks, a heat source, and a suitable trap (usually structural) (Lang and Nadeau, 1984). Oil and gas exploration typically requires extensive geologic mapping and assessment of large areas using geophysical techniques such as gravity, magnetics, and seismic surveys to detect suitable targets. While direct remote sensing detection of hydrocarbons at the surface has been demonstrated using radar data for offshore occurrences (slicks) (MacDonald et al., 1993, 1996; De Beukelaer, 2003) and hyperspectral data for onshore environments (seeps) using hydrocarbon absorption features near 2.3 μm (Ellis et al., 2001), exploration more typically relies on detection and mapping of indirect indicators such as geologic structure, lithology, and detection of specific indirect spectral indicators such as formation of clays, bleaching of iron-rich rocks, and vegetation anomalies (Lang et al., 1984, 1987; van der Meer et al., 2002; Sabins, 1997; Prost, 2001; Khan and Jacobson, 2008). Remote sensing technology is also used to map specific lithologies, to detect geologic structures exposed at the surface, and to help with the siting of geophysical surveys.

Two remote sensing approaches are used for oil and gas exploration (Lang and Nadeau, 1984; Sabins, 1997; Prost, 2001). The first approach, photogeologic interpretation, has been used since the 1930s with data sources as diverse as (initially) black-and-white photography, (more recently) true-color, and color-infrared aerial photographs, synthetic aperture radar (SAR), and panchromatic



Resource Exploration, Figure 1 Left: Reflectance spectra of Fe-bearing secondary minerals and pyrite showing Fe absorptions (Spectra are offset vertically for clarity). Spectra are from the USGS Digital Spectral Library (Clark et al., 1993a) and from Swayze et al. (2000). Right: SWIR (Airborne Visible/Infrared Imaging Spectrometer, AVIRIS) reflectance spectra for selected minerals.

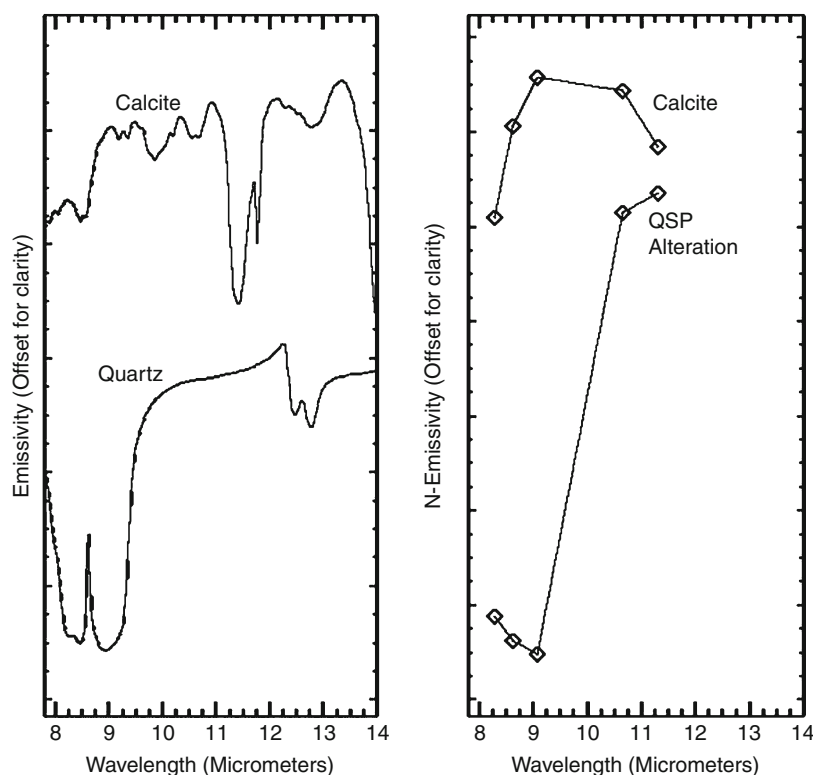
and multispectral satellite imagery (Sabins, 1997). This approach uses the topographic expression of lithology and geologic structure to infer subsurface conditions favorable for oil and gas accumulations (Prost, 2001). The second approach, spectral remote sensing, uses multispectral and/or hyperspectral data to measure spectral properties of materials, thus extracting compositional information about surface materials. Visible and Near-Infrared (VNIR) data (approximately 0.4–1.2 μm) are used to map iron oxide-bearing minerals (hematite, goethite, jarosite) and vegetation distributions (Figure 1, left). Short-Wave-Infrared (SWIR) data (2.0–2.5 μm) can directly map minerals such as clays and carbonates (Figure 1, right). Long-Wave-Infrared (LWIR) data (8.0–14.0 μm) allow detection and quantification of silica (Lyon, 1964; Rowan, 1998) (Figure 2).

Airborne and satellite remote sensing can provide information to assist with resource exploration, assessment, management, and production. Imagery from multispectral satellite sensors such as Landsat and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) gives project managers a synoptic view of producing fields and potential exploration targets/areas. New high-resolution panchromatic and multispectral satellite systems provide spatial resolution better than 1 m, often acting as de facto base maps for

the layout of 2D and 3D seismic surveys, geochemical surveys, and drill sites, as well as logistics planning. Hyperspectral sensors allow detailed mapping of surface mineralogy and vegetation distributions.

Multispectral versus hyperspectral technology

Multispectral imaging (MSI) sensors usually have only a few spectral bands (<20), cover broad spectral regions, and provide synoptic coverage. They do not generally allow direct identification of specific materials, but rather broad groups of materials (e.g., vegetation, clays or carbonates, iron oxides). Examples include Landsat, SPOT, IRS, and ASTER. MSI sensors with high spatial resolution have recently become available (e.g., IKONOS, QuickBird, WorldView). The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is a NASA facility instrument on the Earth Observing System (EOS) TERRA platform that provides Visible/Near-Infrared/Short-Wave-Infrared/Long-Wave-Infrared (VNIR/SWIR/LWIR) Earth observation capabilities in a total of 14 total spectral bands (+one backward-looking band) (Kahle et al., 1991; Fujisada et al., 1998; Yamaguchi et al., 1998; JPL ASTER Website, 2012). ASTER and/or ASTER-simulated data (MODIS/ASTER Airborne Simulator) have been successfully used for



Resource Exploration, Figure 2 Left: Selected LWIR emissivity spectra for the mineral calcite and quartz. Right: Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) LWIR emissivity for pixels containing these materials (QSP Alteration Quartz-Sercite-Pyrite alteration – a predominantly quartz-bearing mineralogy) (From Kruse, 2002).

geologic applications, providing basic mapping capabilities using both the VNIR/SWIR and LWIR spectral ranges (Hook, 1990; Hook et al., 2001; Rowan, 1998; Kruse, 2000, 2002; Rowan and Mars, 2003; Rowan et al., 2003, 2005). The four VNIR bands provide information about iron mineralogy and some rare earth minerals (Rowan and Mars, 2003). The six SWIR bands (Figure 1) allow mapping of molecular vibration absorption features commonly seen in minerals such as carbonates and clays (Hook et al., 2001; Kruse, 2000, 2002). ASTER has collected global multispectral coverage of the Earth between approximately 83° N and 83° S latitudes.

Imaging Spectrometers or “Hyperspectral Imaging (HSI)” sensors, measuring up to hundreds of spectral bands, provide a unique combination of both spatially contiguous spectra and spectrally contiguous images of the Earth’s surface unavailable from other sources (Goetz et al., 1985). HSI data have been used in the geologic community since the early 1980s and represent a mature technology (Goetz et al., 1985; Lang et al., 1987; Pieters and Mustard, 1988; Kruse, 1988, 2012; Crowley, 1993; Boardman and Kruse, 1994; Clark et al., 1996, 1993b; Boardman and Huntington, 1996; Crowley and Zimbelman, 1996; Rowan et al., 1996, 2003, 2004; Cudahy et al., 2001; Kruse et al., 1993, 1999, 2003, 2006). Current airborne HSI sensors provide high

spatial resolution (<1–20 m), high spectral resolution (2–20 nm), and high SNR (>500:1) data for a variety of scientific disciplines.

The Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) represents the current state of the art. AVIRIS, flown by NASA/Jet Propulsion Laboratory (JPL), is a 224-channel imaging spectrometer with approximately 10 nm spectral resolution covering the 0.4–2.5 μm spectral range (Green et al., 1998). The sensor is a whiskbroom system utilizing scanning foreoptics to acquire cross-track data. The IFOV is 1 mrad. Four off-axis double-pass Schmidt spectrometers receive incoming illumination from the foreoptics using optical fibers. Four linear arrays, one for each spectrometer, provide high sensitivity in the 0.4–0.7 μm , 0.7–1.2 μm , 1.2–1.8 μm , and 1.8–2.5 μm regions, respectively. AVIRIS is flown as a research instrument on the NASA ER-2 aircraft at an altitude of approximately 20 km, resulting in approximately 20 m pixels and a 10.5 km swath width. Since 1998, it has also been flown on a Twin Otter aircraft at low altitude, yielding 2–4 m spatial resolution.

The launch of NASA’s EO-1 Hyperion sensor in November 2000 marked the establishment of spaceborne VNIR/SWIR hyperspectral mapping capabilities. Hyperion is a satellite hyperspectral sensor covering the

0.4–2.5 μm spectral range with 242 spectral bands at approximately 10 nm spectral resolution and 30 m spatial resolution from a 705 km orbit (Pearlman et al., 2003). Hyperion is a pushbroom instrument, capturing 256 spectra each with 242 spectral bands over a 7.5 km-wide swath perpendicular to the satellite motion along an up to 160 km path length. The system has two grating spectrometers, one Visible/Near-Infrared (VNIR) spectrometer (approximately 0.4–1.0 μm) and one Short-Wave-Infrared (SWIR) spectrometer (approximately 0.9–2.5 μm). Key Hyperion characteristics are discussed further in Green et al., 2003. Hyperion data are available for purchase from the US Geological Survey (USGS EO-1 Website, 2012). Thousands of Hyperion scenes have been acquired for a variety of disciplines. The EO-1 Science Validation Team has evaluated and validated the instrument. Selected results have been presented at team meetings (NASA Goddard EO-1 Website, 2012) and also published at various venues (Asner and Green, 2001; Hubbard and Crowley, 2001; Kruse et al., 2003). Also see Ungar et al. (2003) for a summary along with associated papers.

Mineral exploration examples

Case history: mineral mapping using hyperspectral remote sensing at Mt. Fitton, Australia

HyMap (Integrated Spectronics, Sydney, Australia) is a commercially available hyperspectral sensor that provides combined high spatial resolution (2–10 m), high spectral resolution (12–16 nm), and high signal-to-noise (>500:1) performance. Figure 3 shows images and spectra of the Mt. Fitton area, South Australia, acquired by the HyMap scanner. The local geology in this area is comprised of a series of folded unconformable Adelaidean (Proterozoic) sediments syn- and post-intruded by Proterozoic diapirs, and Palaeozoic granitic and granodioritic intrusives (Reston and Cocks, 1998). Sedimentary facies show a full range of argillaceous, arenaceous, and carbonaceous compositions and textures. Palaeozoic tectonism is characterized by E/ENE trending asymmetric folds with steeply dipping axial traces. Synform features are strongly developed within the area. The HyMap data shown in Figure 3 cover an approximately 2.6×4 km area at 5 m spatial resolution in 128 narrow (10–20 nm-wide) spectral bands. Analysis of these HyMap data without a priori information and using only the 2.0–2.5 μm range indicates the presence of over 15 distinct minerals associated with primary lithology and alteration. Figure 3 shows a selection of spectra for these minerals and the mapping results.

Case history: mapping silica, alunite, kaolinite at an active hot springs system

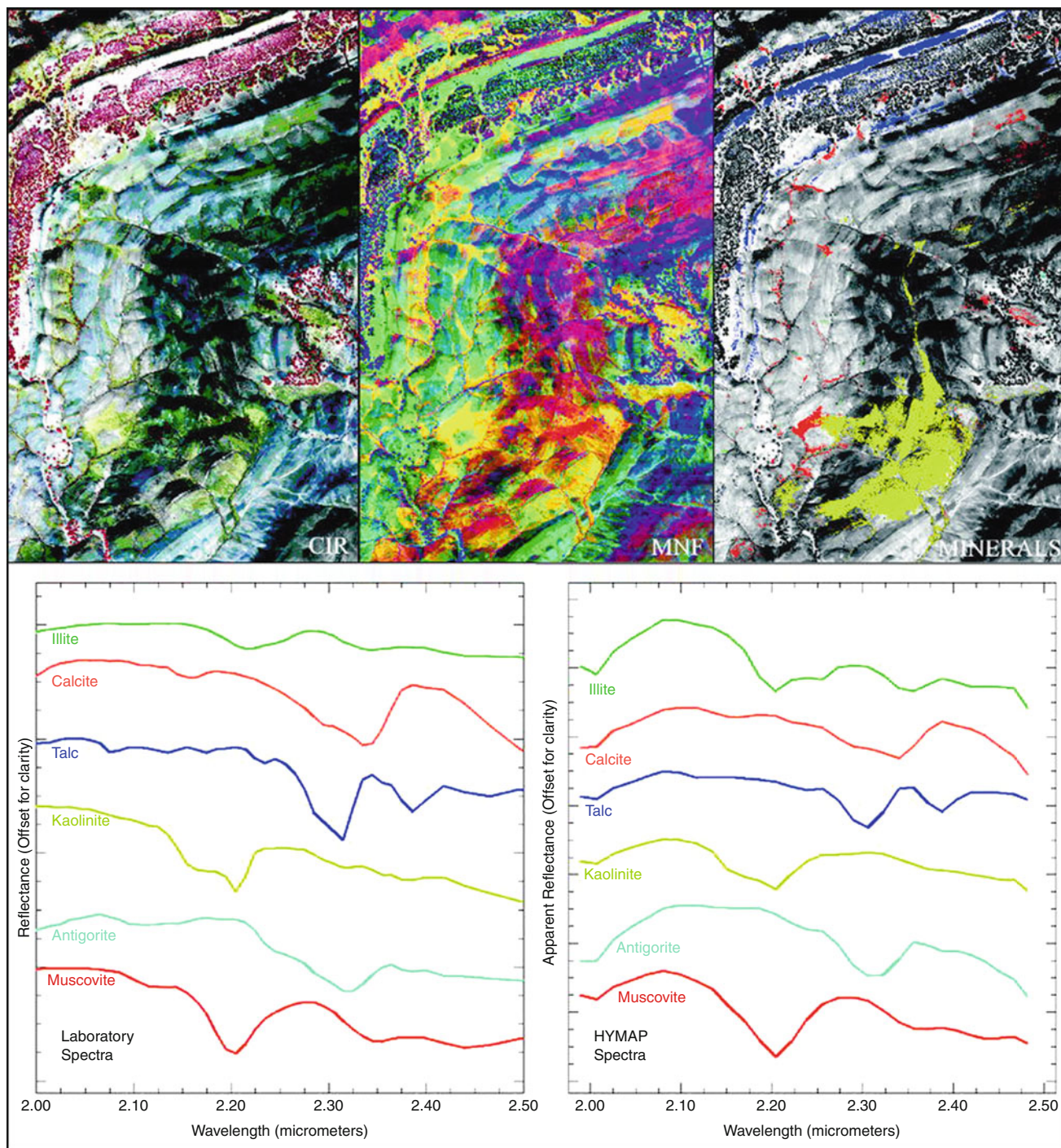
Study of the nature and spatial distribution of specific hydrothermal alteration minerals at hot springs using hyperspectral remote sensing provides insight into system geochemistry, and to the occurrence and characteristics of hot springs systems in the fossil record, potentially leading to new and/or improved exploration methods for ore

deposits. The Steamboat Springs hydrothermal system is described as a present-day equivalent of epithermal gold–silver deposits (White, 1955, 1967). This hydrothermal system, located just south of Reno, Nevada (Figure 4), is associated with four rhyolite domes, and thermal activity has probably been continuous for at least the past 0.1 m.y (Silberman et al., 1979). Numerous wells have been drilled at Steamboat for geothermal energy and to obtain hot water for local resort facilities. Early exploration wells ranged from 218 to 558 m with a maximum measured temperature of 186 °C (White, 1968, 1981). Geothermal resource exploitation for power production commenced in the early 1980s, with the first production facility producing 14.4 MW starting in 1988. A new well brought into production in July 2000 had a high temperature of 248°C (Nevada Bureau of Mines, 2008). The principal surface mineralogy reported at Steamboat consists of chalcedonic sinter deposits (Figures 4 and 5). Prior to 1988, dark siliceous muds were also being deposited in the active springs; however, this activity ceased after start of power production. Acid-leached opaline residues, kaolinite, and alunite occur in solfatarically altered granodiorite and basaltic andesite in the western part of the area (Figures 4 and 5) (Sigvaldason and White, 1962; White et al., 1964; Schoen and White, 1967; Schoen et al., 1974). Significant concentrations of precious metals and related pathfinder elements occur in the Steamboat Springs sinter deposits, as chemical sediments in spring vents and as veins at depth (White, 1981; Nevada Bureau of Mines, 2008). Gold was detected at the 1–2 ppm level along with anomalous Ag and As concentrations in analysis of samples from several drill holes, and small amounts of Hg has been mined from the Mercury mine at Steamboat (White et al., 1992). Deep drilling at Steamboat shows vein and alteration patterns that are indistinguishable from those of many epithermal ore deposits, containing adularia, illite, montmorillonite, and chlorite-group minerals as well as kaolinite, chalcedony, calcite, and quartz. Both stibnite and cinnabar are present near the surface; however, ore-grade concentrations of metals appear to be absent both in the near-surface deposits and in the veins at depth.

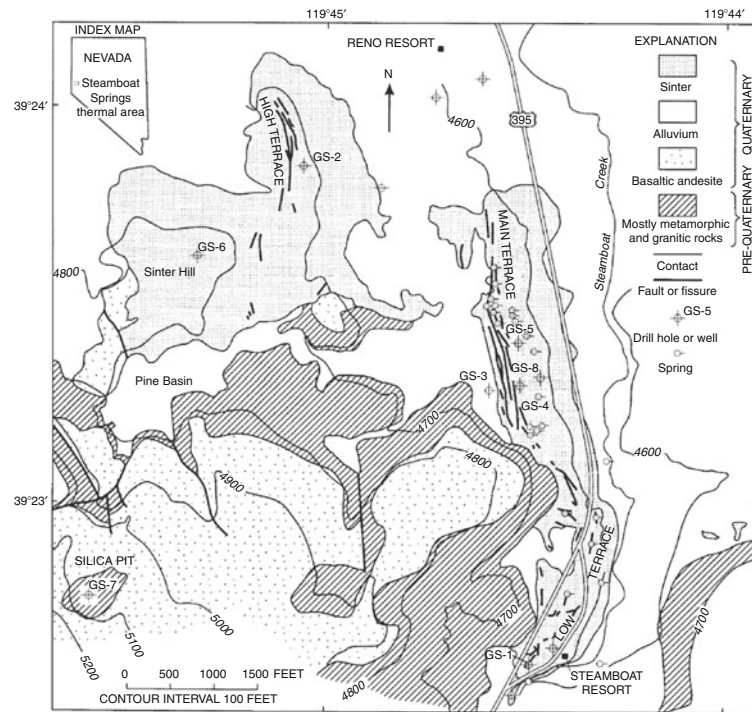
AVIRIS data were acquired during July 1995 for the Steamboat Springs area and during October 1998 as part of the JPL AVIRIS low-altitude test program. AVIRIS data from both flights were calibrated to apparent reflectance using the ATREM method (Gao and Goetz, 1990; Kruse, 2004; Gao et al., 2009). Data were then analyzed using standardized procedures (Boardman et al., 1995; Kruse et al., 2003; Boardman and Kruse, 2011). Figure 6 shows an image, endmember spectra, and mineral map for the 1998 data (~2.2 m pixels).

Case history: gold exploration using hyperspectral remote sensing

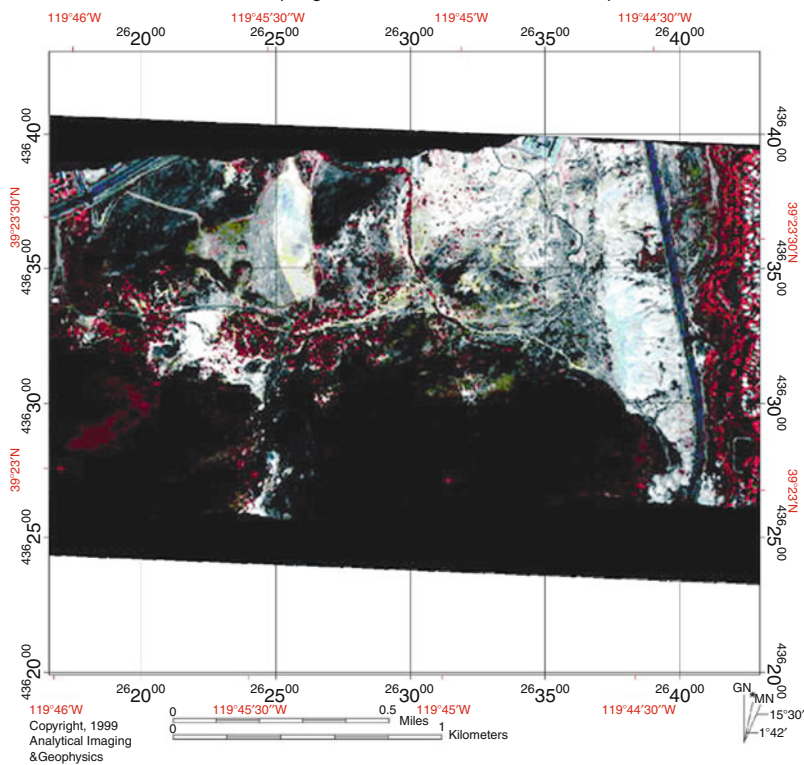
Numerous gold deposits occur in Eureka County, North Central Nevada, along a regional mineral zone known as the Cortez or Battle Mountain-Eureka Trend (Figure 7).



Resource Exploration, Figure 3 HyMap data of the Mt. Fitton area, South Australia (Data courtesy of Integrated Spectronics Pty Ltd, acquired March 1998). The *left* image is a Color-Infrared Composite (CIR) Image of HyMap bands 30, 17, and 10 (0.863, 0.665, 0.557 μm) (RGB). The *center* image shows bands 1, 2, and 3 (RGB) of a Minimum Noise Fraction Transform (MNF) of selected VNIR bands. The *left* plot shows laboratory reflectance spectra from the US Geological Survey spectral library for selected minerals. The *right* plot shows average apparent reflectance spectra for selected minerals extracted from the HyMap hyperspectral data identified using visual comparison to the USGS spectral library. The *right* image shows color overlays of these selected mineral occurrences on a grayscale image of HyMap Band 17 (0.665 μm). Colors in the *right* image correspond to those used in the spectral plots. Additional information about HyMap is available at <http://www.intspec.com>.



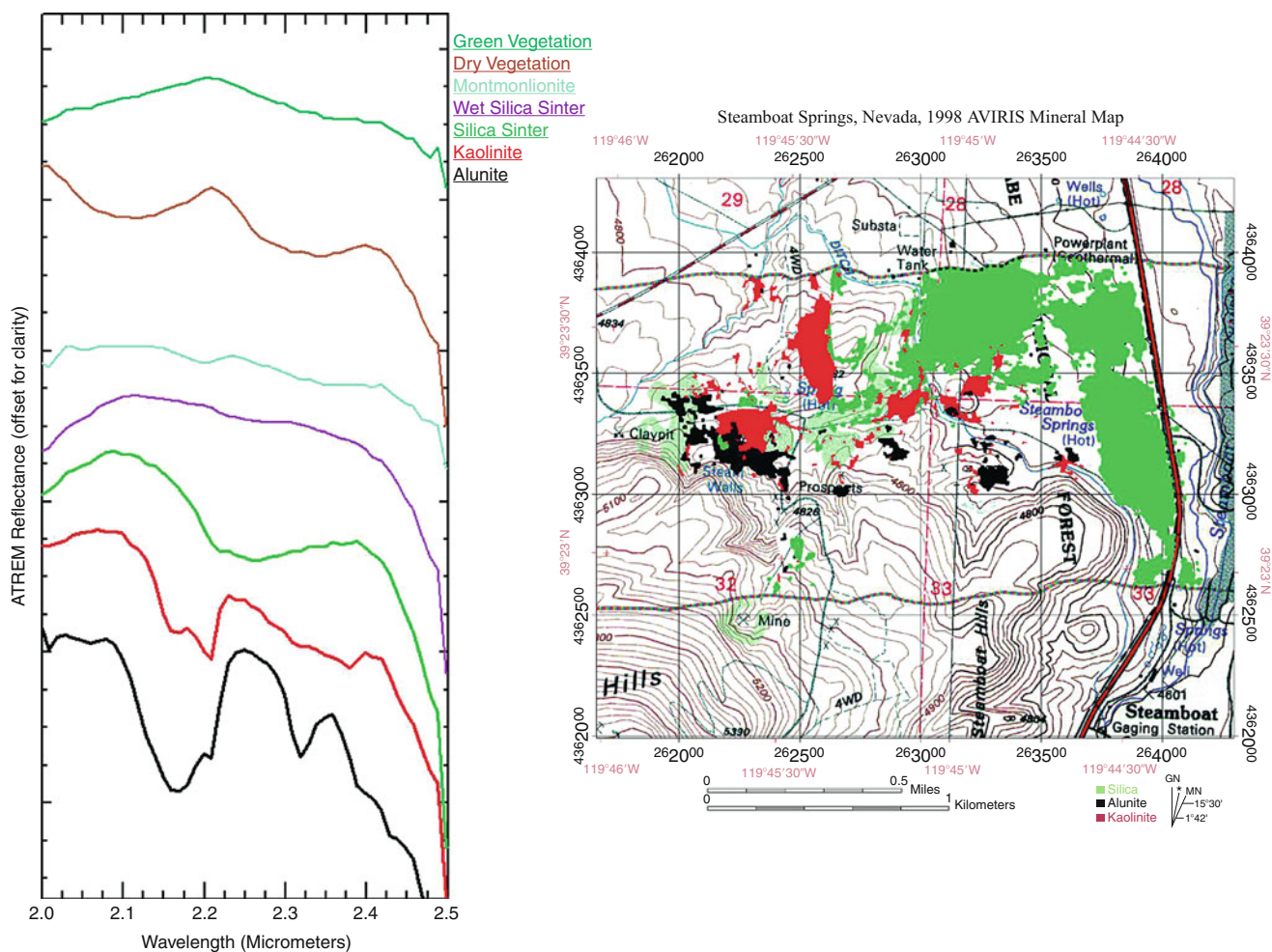
Steamboat Springs, Nevada, 1998 AVIRIS CIR Composite



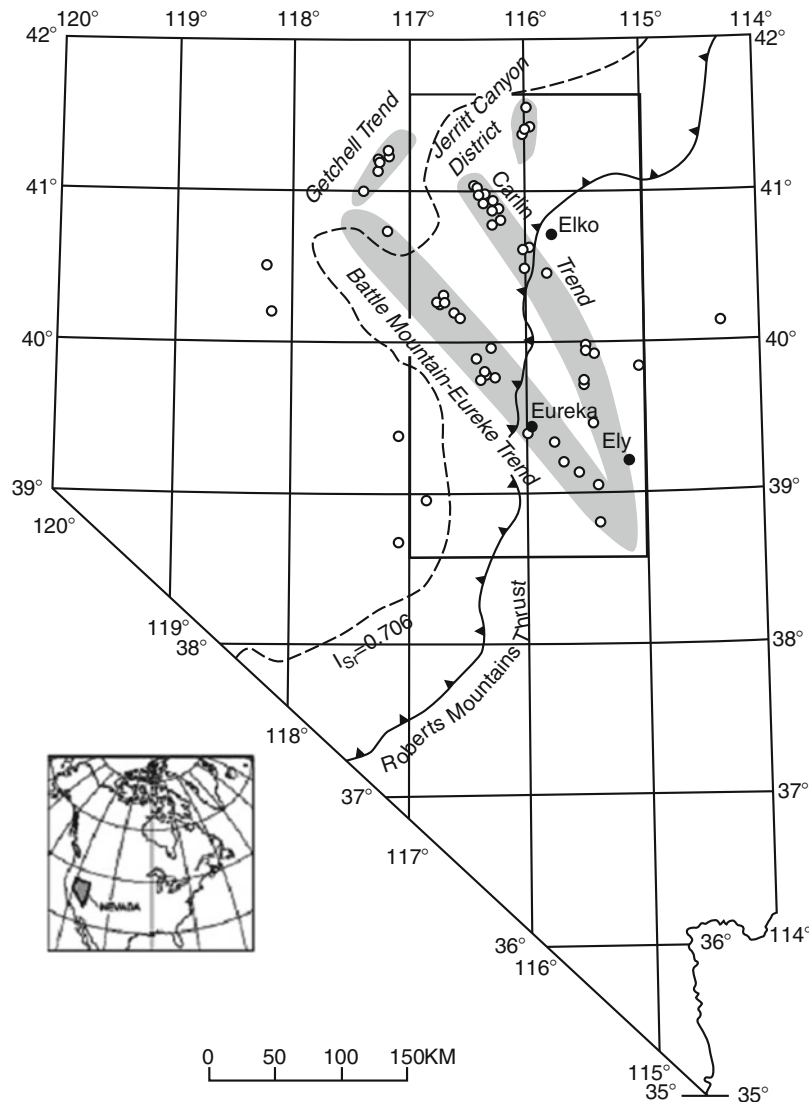
Resource Exploration, Figure 4 Top: Location and geology of Steamboat Springs, Nevada (From Silberman et al., 1979). Bottom: AVIRIS Color-Infrared Composite Image.



Resource Exploration, Figure 5 Ground-level photographs at Steamboat Springs, Nevada, circa 1987. *Left photo* shows silica sinter surface. *Right photo* shows acid-sulfate area.



Resource Exploration, Figure 6 *Left*: AVIRIS endmember mineral spectra. *Right*: AVIRIS mineral map of selected minerals overlain on topographic map (red kaolinite, green silica, black alunite).

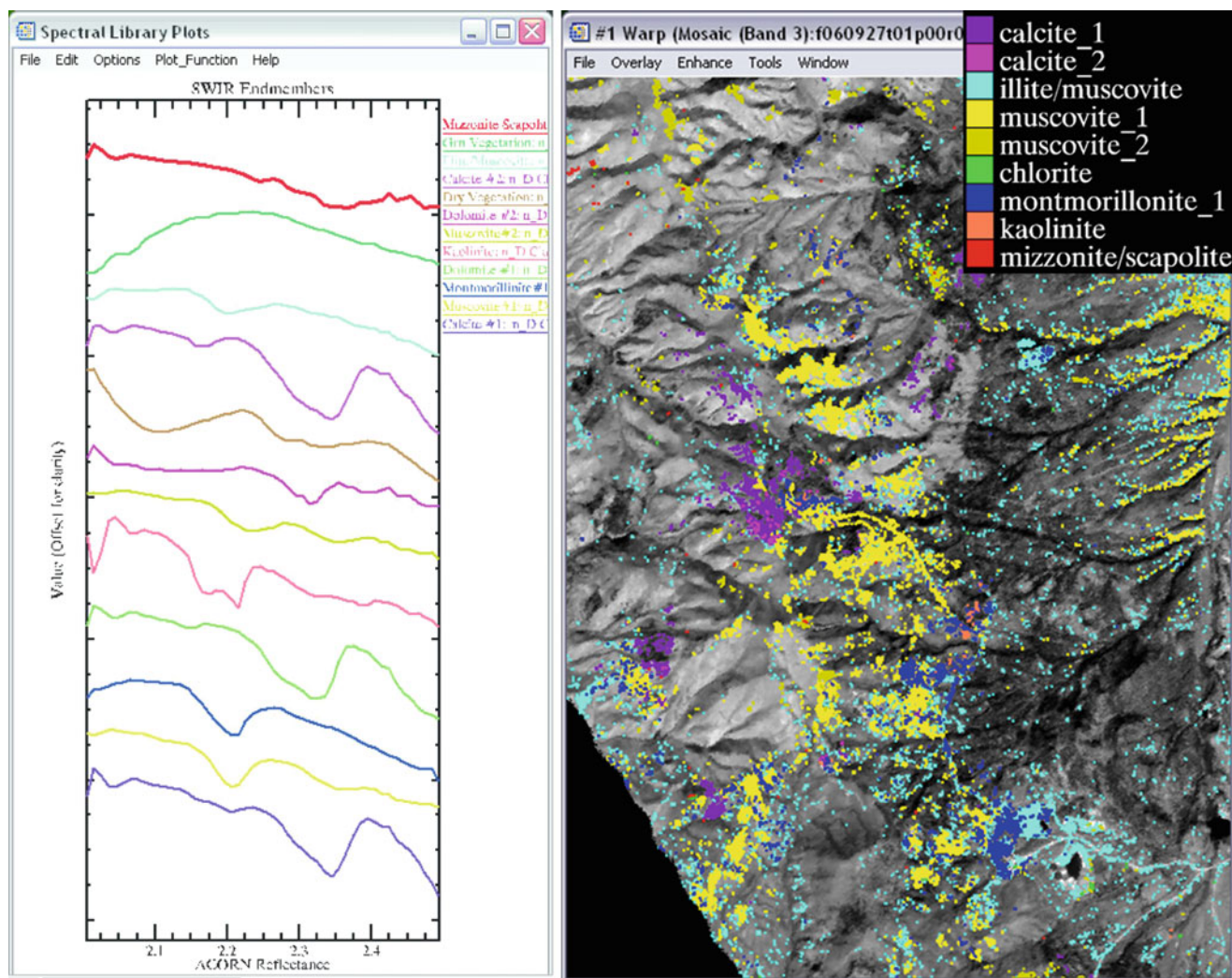


Resource Exploration, Figure 7 Battle Mountain trend location map.

Disseminated, sedimentary-hosted gold deposits occur along this trend associated with Palaeozoic regressive-transgressive, continental shelf facies (Broili et al., 1988). This style of gold mineralization is also known as the Carlin-style deposit model, where host rocks include siltstone, limestone, marls, and other Palaeozoic nearshore units. Generally, gold deposits are associated with fluid movement along feeder structures. Hydrothermal fluids flowing upsection and updip decalcify and alter (or “sand”) host units, which typically are silty or shaley carbonate units. This initial hydrothermal phase prepares the host rock by increasing porosity and permeability, and gold is deposited during a second alteration stage as microscopic particles within the altered silty calcareous units. Silica and iron mixtures can solidify near or adjacent

to gold deposits forming jasperoids, often manifested as resistant red-color ridges within carbonates at the surface. Generally, associated alteration minerals are limited to silica, skarn minerals, pyrite, iron oxides (jasperoids), and limited argillic alteration minerals.

Recent exploration in this area by Perry Remote Sensing (PRS)/Horizon GeoImaging LLC using remote sensing data consisted of analysis of ASTER data to predict mineral components expected for the area and detailed mineral analysis using AVIRIS. The entire Battle Mountain-Eureka mineral trend was the focus of initial mineral mapping using ASTER without ground truth information. Fieldwork during April 2006, including spectral measurements using a field-portable spectrometer, demonstrated the accuracy of the ASTER mapping



Resource Exploration, Figure 8 Left: AVIRIS SWIR mineral endmembers. Plotted from top to bottom: Mizzonite/Scapolite, Green Vegetation, Illite/Muscovite, Calcite_2, Dry Vegetation, Dolomite_2, Muscovite_2, Kaolinite, Dolomite_2, Montmorillonite_1, Muscovite_1, Calcite_1. Right: AVIRIS mineral mapping results. Red triangles represent areas where field investigation and spectrometer measurements verify AVIRIS mapping. AVIRIS image width is approximately 10 km.

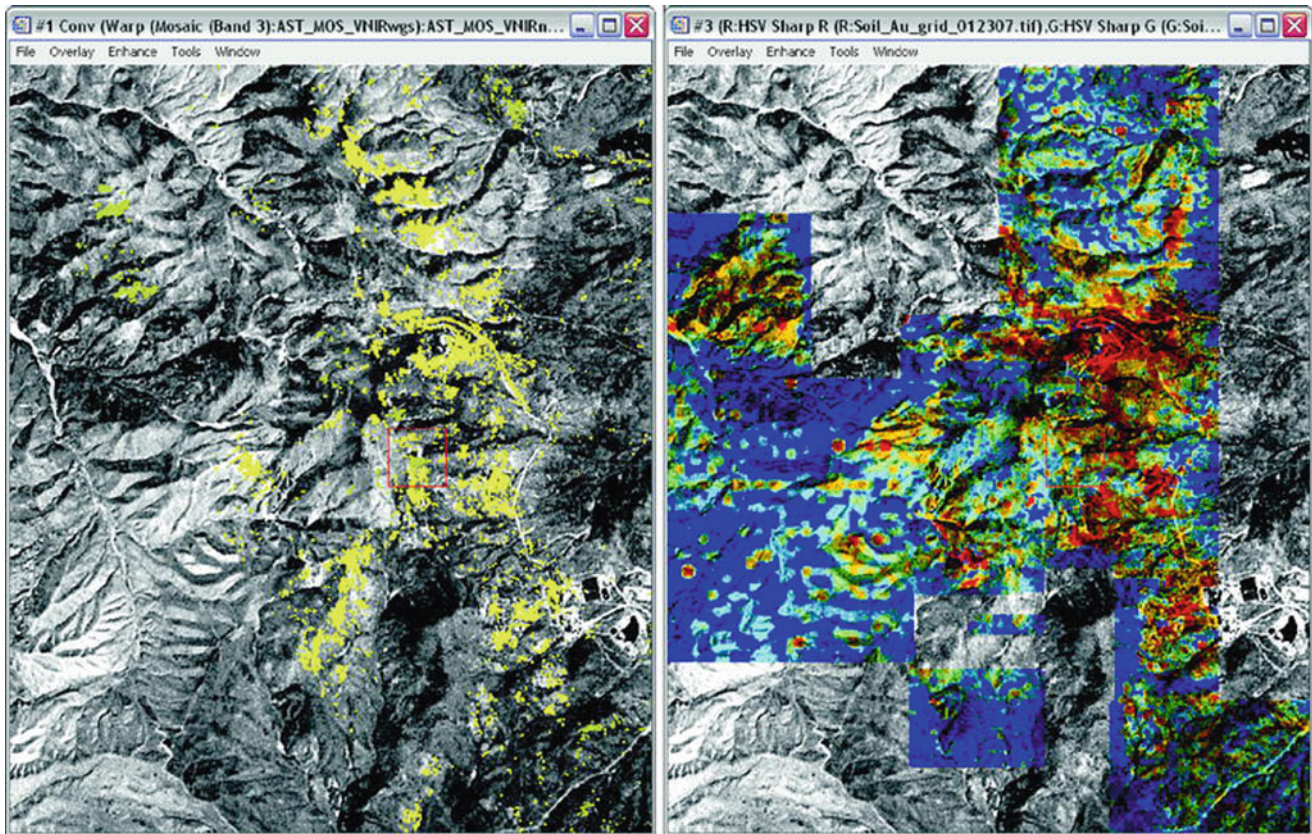
for muscovite/sericite and carbonates. Low-altitude AVIRIS imagery (3.5 m resolution) was acquired along the entire Battle Mountain-Eureka Trend as two overlapping flight lines collected on September 27, 2006. Each AVIRIS flight line was processed and analyzed separately using VNIR and SWIR spectral ranges and standardized procedures (Boardman et al., 1995; Kruse et al., 2003; Boardman and Kruse, 2011). AVIRIS SWIR processing and analysis revealed 15 mineral endmembers including several varieties of carbonates and muscovite (Figure 8). Together, the muscovite and muscovite/illite endmember classes cover the greatest extent of the AVIRIS survey. Illite/muscovite appears to be the primary mineral component of the pediment gravels and broad alluvial basins. Of particular interest, muscovite

#1 spatially coincides with a Jurassic intrusion near a known gold district as well as a large molybdenum deposit near Eureka called Mount Hope. This same muscovite class is mapped in other areas and may reflect unmapped alteration associated with intrusive activity. Comparison to geochemical results shows a strong correspondence between AVIRIS-mapped muscovite #1 and gold concentrations (Figure 9).

Oil and gas exploration example

Case history: Patrick Draw, Wyoming

The following information is mostly extracted and modified from the section describing the Patrick Draw test site in the Joint NASA/GeoSat Test Case Final Report



Resource Exploration, Figure 9 Left: AVIRIS muscovite #1 image map for a subarea, Right: Gold concentration in soils for the same area color coded from dark blue, to cyan, to green, to yellow, to red from low to high concentration. Note correspondence of high gold values (red) with muscovite #1 mineral mapping.

(Lang et al., 1984). Patrick Draw is a well-known hydrocarbon occurrence (Cox, 1962; McCubbin and Brady, 1963; Weimer, 1966), and the test site has been used for a variety of remote sensing tests for hydrocarbon exploration (Richers et al., 1982, 1986; Lang et al., 1984; Scott et al., 1989; Khan and Jacobson, 2008 and others). The site is located in South Central Wyoming approximately 55 km east of the city of Rock Springs, Wyoming (Figure 10).

The Patrick Draw field consists of oil and gas in stratigraphic traps located on the eastern flank of the Rock Springs Uplift (Figure 11). It was first discovered in 1959 based on observation of the “pinch out” of known gas-producing sandstones against the Rock Springs Uplift. The principal productive interval consists of permeable and porous sandstone bodies at the top of the Upper Cretaceous Almond Formation that grade up dip on the west into impermeable shale and sandstones (Weimer, 1966). Extensive work in support of the NASA/GeoSat Test Case Project demonstrated suggestive correlations between remote sensing data and suspected gas microseepage (Richers et al., 1982, 1986; Lang et al., 1984). While designed as a “direct detection” experiment, this case history demonstrates remote sensing

approaches and methods that could be applied to exploration for typical oil and gas occurrences without direct indicators. Multispectral remote sensing (Landsat MSS and NS-001 Thematic Mapper Simulator) was used to map evidence of structural traps. NS-001 photointerpretation at a scale of 1:48,000 showed broad folds that were not detected in previously available subsurface structure maps of the site. Lineament analysis of Landsat MSS, topography, NS-001, and radar images identified stress orientations associated with jointing, faulting, and folding that when compared with known structures helped assess the timing and orientation of structural disturbance. NS-001 Thematic Mapper Simulator imagery provided the ability to recognize and trace stratigraphic units at the surface, essential for the preparation of isopach, facies, and other stratigraphic maps and to help locate potential stratigraphic traps and to develop basin evolution models. The NS-001 data also allowed recognition of a previously unmapped lithostratigraphic unit and tracing of individual discontinuous beds that were difficult (or impossible) to trace using standard field techniques. Soils and vegetation mapping using the NS-001 discriminated between soil units, providing detail similar to existing BLM soil maps, which required years of

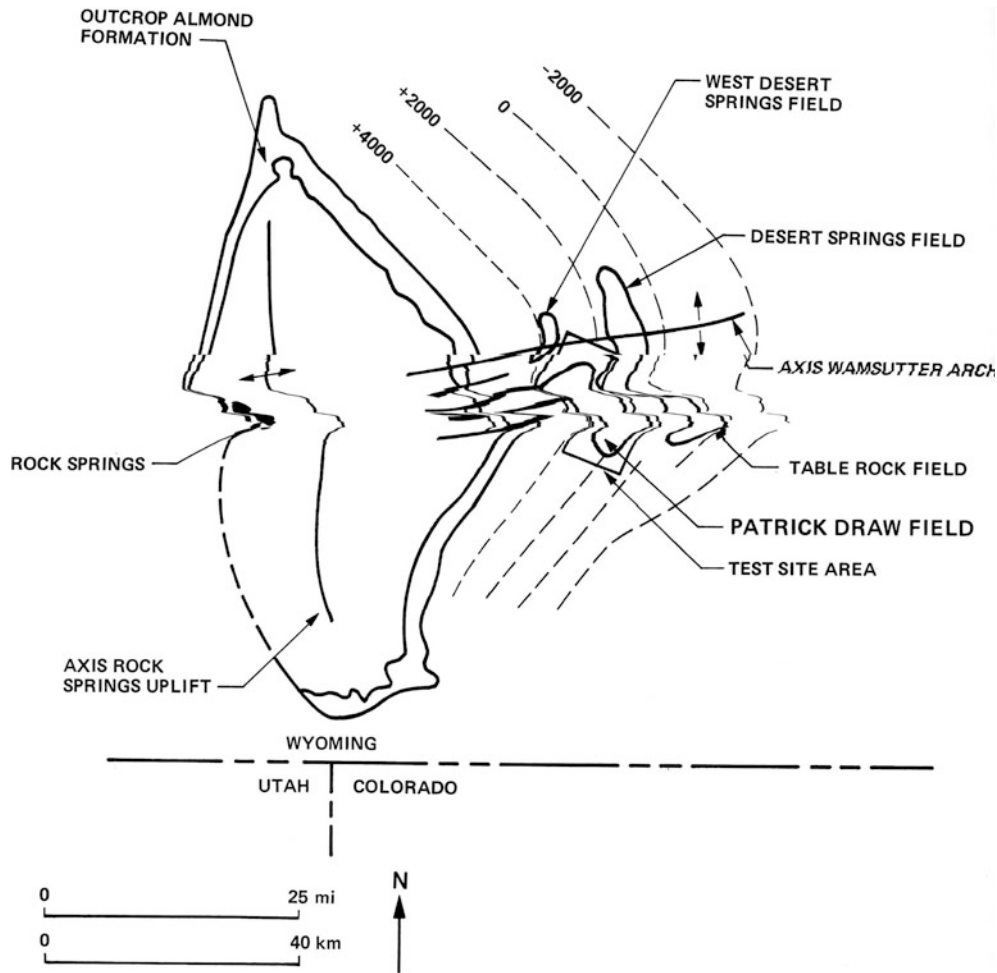


Resource Exploration, Figure 10 Map showing the location of the Patrick Draw test site (From Lang et al., 1984).

fieldwork to prepare. Vegetation mapping using the multispectral data showed detail not available using other means (including standard aerial photography) and correlated well with geochemical investigations, including soil–gas surveys (Figure 12). A previously unknown area of abnormal, stunted sage corresponding approximately to the center of the gas cap of the underlying Patrick Draw reservoir was discovered using the NS-001 data. The area of stunted sage occurrence also shows anomalously high soil

hydrocarbons, the highest soil helium concentration, and anomalously high concentrations of titanium, vanadium, copper, and zinc, along with low concentrations of calcium and magnesium.

Subsequent studies comparing Landsat Thematic Mapper data to the soil–gas anomalies at Patrick Draw included a more detailed soil–gas survey on and around the area’s producing fields (Richers et al., 1982, 1986). The results of this study agree with the NASA/GeoSat



Resource Exploration, Figure 11 Regional subsurface structural contour map showing Patrick Draw and adjacent fields. Contours (in feet) on the top of the Almond Formation; datum, sea level (From Lang et al., 1984).

assessment that the faults and fractures visible as linear features on satellite and aircraft imagery provide paths for active microseepage. Additional soil–gas measurements over the known vegetation anomaly revealed a much wider area of anomalously high free soil–gas values and fluorescence than was previously reported.

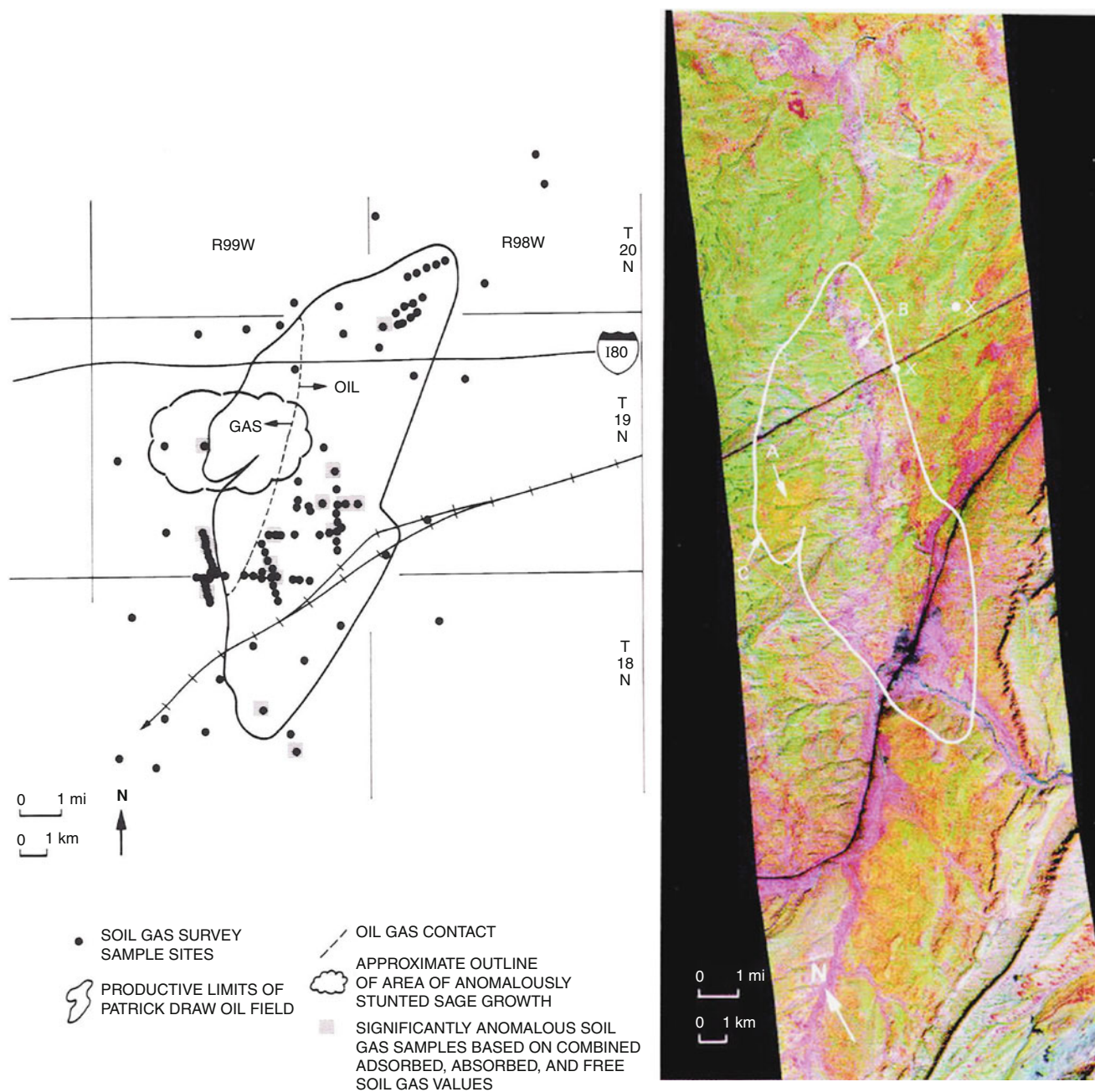
Scott et al. (1989) point out that the stunted sage anomaly at Patrick Draw may in fact be related to increased soil exposure rather than sage morphological changes. While they also indicate that they do not see a definitive link between the tonal anomaly and gas seepage, the GeoSat Test Case and Richers's evidence seem clear.

More recently, hyperspectral remote sensing data supported by spectral and geochemical measurements of surface soils have been used to locate and characterize alteration mineralogy at Patrick Draw (Khan and Jacobson, 2008). Mineralogical, geochemical, and carbon isotope data support the presence of hydrocarbon microseepage. Specific surface changes associated with

the seepage include decreased feldspar content (determined by XRD), increased clay concentrations (from spectral reflectance measurements), and carbon-13 values consistent with hydrocarbon occurrences. Hyperion data (Ungar et al., 2003) were used to map alteration associated with hydrocarbon microseepages in the Patrick Draw area. A northeast–southwest trending anomalous zone following a defined stratigraphic boundary appeared to be one migrational pathway for hydrocarbon seepage (Figure 13). Alteration mineralogy was also associated with the known vegetation anomaly.

Conclusions

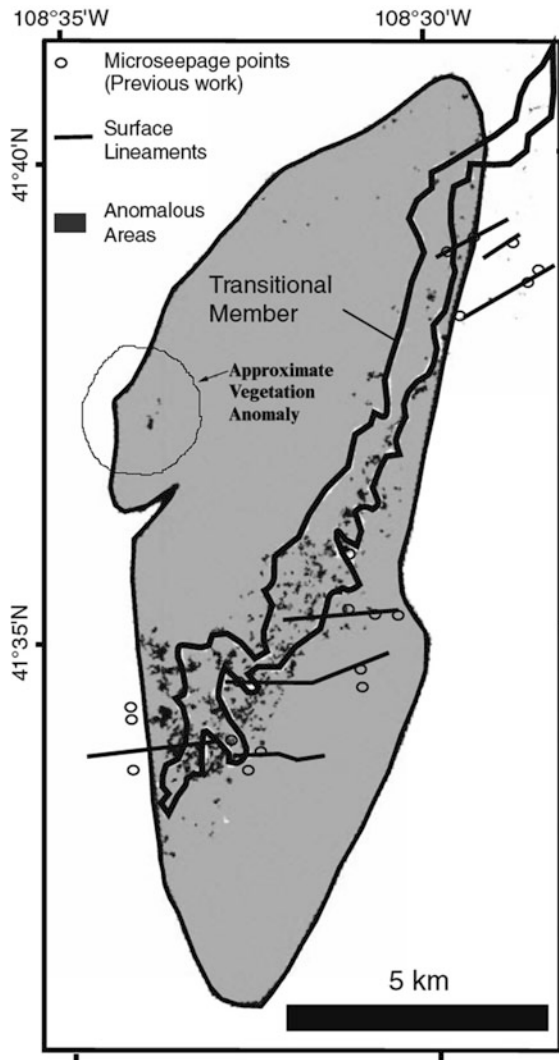
Remote sensing technology provides a unique means of exploring for nonrenewable resources at the Earth's surface and monitoring their development and exploitation. Multispectral satellite imagery provides synoptic, repetitive coverage of most locations and allows both



Resource Exploration, Figure 12 Left: Soil-gas anomaly significance map showing combined results of adsorbed, absorbed, and free soil-gas (Lang et al., 1984, Figure 11–22). Right: NS-001 principle components image (Lang et al., 1984, Figure 11–75). PCs 1, 2, 3 are displayed as RGB. Location “A” shows the total anomaly (a yellow-toned area in a region of green) coinciding with the areal distribution of anomalously stunted sage.

geomorphic and spectral analysis. Radar (SAR) data are also used for mapping geologic structure and for direct detection of seeps in oil exploration. Airborne hyperspectral data (and currently limited satellite data) allow detailed mineralogical mapping important to detection and characterization of indirect indicators of

both minerals and oil and gas occurrences. Case histories described here highlight a few of the methods and applications of remote sensing data to resource exploration. The bibliography provided acts to direct readers to further details, extended case histories, and additional resources.



Resource Exploration, Figure 13 Patrick Draw area, showing previously determined points of microseepage, associated lineaments, anomalous areas (in black), outline of transitional member of Fort Union Formation, and approximate vegetation anomaly extent (From Khan and Jacobson, 2008).

Bibliography

- Asner, G. P., and Green, R. O., 2001. Imaging spectroscopy measures desertification in the Southwest U.S. and Argentina. *Eos. Trans. AGU*, **80**, 601–605, doi:10.1029/01EO00346. <http://dx.doi.org/10.1029/01EO00346> (September 2012).
- Boardman, J. W., and Huntington, J. H., 1996. Mineral mapping with 1995 AVIRIS data. In *Summaries of the 6th Annual JPL Airborne Earth Science Workshop, JPL Pub. 96-4*, AVIRIS Workshop, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, Vol. 1, pp. 9–11.
- Boardman, J. W., and Kruse, F. A., 1994. Automated spectral analysis: A geologic example using AVIRIS data, north Grapevine Mountains, Nevada. In *Proceedings, Tenth Thematic Conference on Geologic Remote Sensing, Environmental Research Institute of Michigan*, Ann Arbor, MI, pp. I-407–I-418.
- Boardman, J. W., and F. A. Kruse, 2011. Analysis of Imaging Spectrometer Data Using N-Dimensional Geometry and A Mixture-Tuned Matched Filtering (MTMF) Approach, *Transactions on Geoscience and Remote Sensing (TGARS), Special Issue on Spectral Unmixing of Remotely Sensed Data*, **49**(11), 4138–4152.
- Boardman, J. W., Kruse, F. A., and Green, R. O., 1995. Mapping target signatures via partial unmixing of AVIRIS data. In *Summaries, Fifth JPL Airborne Earth Science Workshop, JPL Publication 95-1*, Vol. 1, pp. 23–26.
- Broili, C., French, G. M., Shaddrick, D. R., and Weaver, R. R., 1988. Geology and gold mineralization of the Gold Bar deposit, Eureka county, Nevada. In *Bulk Mineable Precious Metal Deposits of the Western United States, GSN Symposium Proceedings*, pp. 57–72.
- Clark, R. N., Swayze, G. A., Gallagher, A., King, T. V. V., and Calvin, W. M., 1993a. *The U. S. Geological Survey Digital Spectral Library: Version 1: 0.2 to 3.0 μm*. Washington, DC: U.S. Government Printing Office. U. S. Geological Survey, Open File Report 93–592, p. 1340. <http://speclab.cr.usgs.gov> (September 2012).
- Clark, R. N., Swayze, G. A., and Gallagher, A., 1993b. *Mapping Minerals with Imaging Spectroscopy*. Washington, DC: U.S. Government Printing Office. U.S. Geological Survey Bulletin 2039B, pp. 141–150.
- Clark, R. N., Swayze, G. A., Rowan, L. C., Livo, K. E., and Watson, K., 1996. Mapping surficial geology, vegetation communities, and environmental materials in our national parks: The USGS imaging spectroscopy integrated geology, ecosystems, and environmental mapping project. In *Summaries of the 6th Annual JPL Airborne Earth Science Workshop, JPL Pub. 96-4*. AVIRIS Workshop, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, Vol. 1, pp. 55–56.
- Cox, J. E., 1962. *Patrick Draw field and Adjacent Areas, Sweetwater County, Wyoming*. Billings Geol. Soc. Paper No. 1, pp. 1–17, Montana Geological Society (2010).
- Crowley, J. K., 1993. Mapping playa evaporite mineral with AVIRIS data: a first report from Death Valley, California. *Remote Sensing of Environment*, **44**(2–3), 337–356.
- Crowley, J. K., and Zimbelman, D. R., 1996. Mapping hydrothermally altered rock on Mount Rainier, Washington, DC: Application of AVIRIS data to volcanic hazard assessments. In *Summaries of the 6th Annual JPL Airborne Earth Science Workshop, JPL Pub. 96-4*. AVIRIS Workshop, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, Vol. 1, pp. 63–66.
- Cudahy, T. J., J. Wilson, R. Hewson, K. Okada, P. Linton, P. Harris, M. Sears, and J. A. Hackwell, 2001. Mapping Porphyry-Skarn Alteration at Yerington, Nevada, Using Airborne Hyperspectral VNIR-SWIR-TIR Imaging Data, in *Proceedings IGARSS Geoscience and Remote Sensing International Symposium*, **2**, 631–633.
- De Beukelaer, S. M., 2003. *Remote Sensing Analysis of Natural Oil and Gas Seeps on The Continental Slope of The Northern Gulf of Mexico*. Unpublished PhD thesis, Texas, Texas A&M University, 117 p. (<http://txspace.tamu.edu/bitstream/handle/1969.1/1164/etd-tamu-2003B-2003070315-De%20B-1.pdf?sequence=1>) (September 2012).
- Ellis, J. M., Davis, H. H., and Zamudio, J. A., 2001. Exploring for onshore oil seeps with hyperspectral imaging. *Oil and Gas Journal*, **99**(37), 49–58.
- Farrand, W. H., 1997. Identification and mapping of ferric oxide and oxyhydroxide minerals in imaging spectrometer data of Summitville, Colorado, U.S.A., and the surrounding San Juan Mountains. *International Journal of Remote Sensing*, **18**(7), 1543–1552.

- Fujisada, H., Sakuma, F., Ono, A., and Kudoh, M., 1998. Design and preflight performance of ASTER instrument protoflight model. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6), 1152–1160, doi:10.1109/36.701022. <http://dx.doi.org/10.1109/36.701022> (September 2012).
- Gao, B., and Goetz, A. F. H., 1990. Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. *Journal of Geophysical Research*, **95**(D4), 3549–3564.
- Gao, B., Montes, M. J., Davis, C. O., and Goetz, A. F. H., 2009. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sensing of Environment*, **113**, 517–524.
- Goetz, A. F. H., Vane, G., Solomon, J. E., and Rock, B. N., 1985. Imaging spectrometry for earth remote sensing. *Science*, **228**, 1147–1153.
- Green, R. O., Eastwood, M. L., and Sarture, C. M., 1998. Imaging spectroscopy and the airborne visible infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment*, **65**(3), 227–248, doi:10.1016/S0034-4257(98)00064-9. [http://dx.doi.org/10.1016/S0034-4257\(98\)00064-9](http://dx.doi.org/10.1016/S0034-4257(98)00064-9) doi:10.1016/S0034-4257%2898%2900064-9 (September 2012).
- Green, R. O., Chrien, T. G., and Pavri, B., 2003. On-orbit determination of the radiometric and spectral calibration of Hyperion using ground, atmospheric and AVIRIS underflight measurements. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6), 1194–1203, doi:10.1109/TGRS.2003.813204. <http://dx.doi.org/10.1109/TGRS.2003.813204> (September 2012).
- Hook, S. J., 1990. The combined use of multispectral remotely sensed data from the short wave infrared (SWIR) and thermal infrared (TIR) for lithological mapping and mineral exploration. In *Proceedings, Fifth Australasian Remote Sensing Conference*, Perth, Western Australia, 1, pp. 371–380.
- Hook, S. J., Myers, J. J., Thome, K. J., Fitzgerald, M., and Kahle, A. B., 2001. The MODIS/ASTER airborne simulator (MASTER) – a new instrument for earth science studies. *Remote Sensing of Environment*, **76**, 93–102, doi:10.1016/S0034-4257(00)00195-4. [http://dx.doi.org/10.1016/S0034-4257\(00\)00195-4](http://dx.doi.org/10.1016/S0034-4257(00)00195-4) doi:10.1016/S0034-4257%2800%2900195-4 (September 2012).
- Hubbard, B. E., and Crowley, J. K., 2001. Alteration mineral mapping in the Central Andes using Hyperion, ALI and ASTER. *Geological Society of America*, **33**(6), A-319. Abstracts Programs, https://gsa.confex.com/gsa/2001AM/finalprogram/abstract_22837.htm (September 2012).
- JPL ASTER Website, 2012. <http://asterweb.jpl.nasa.gov/instrument.asp> (September 2012).
- Kahle, A. B., Palluconi, F. D., Hook, S. J., Realmuto, V. J., and Bothwell, G., 1991. The advanced spaceborne thermal emission and reflectance radiometer (ASTER). *International Journal of Imaging Systems and Technology*, doi:10.1002/ima.1850030210.
- Khan, S. D., and Jacobson, S., 2008. Remote sensing and geochemistry for detecting hydrocarbon microseepages. *Geological Society of America*, **120**(1–2), 96–105, doi:10.1130/0016-7606(2008)120[96:RSAGFD]2.0.CO;2.
- Kruse, F. A., 1988. Use of Airborne Imaging Spectrometer data to map minerals associated with hydrothermally altered rocks in the northern Grapevine Mountains, Nevada and California. *Remote Sensing of Environment*, **24**(1), 31–51.
- Kruse, F. A., 1996. Mineral mapping for environmental hazards assessment using AVIRIS data, Leadville, Colorado. In *Proceedings, 11th Thematic Conference, Applied Geologic Remote Sensing*, February, 27–29, 1996. Ann Arbor, MI: Environmental Research Institute of Michigan (ERIM), pp. II-526–II-533.
- Kruse, F. A., 1999. Mapping hot spring deposits with AVIRIS at steamboat springs, Nevada. In *Proceedings of the 8th JPL Airborne Earth Science Workshop: Jet Propulsion Laboratory Publication 99–17*, Pasadena, CA, pp. 239–246.
- Kruse, F. A., 2000. Mapping active and fossil hot springs systems using AVIRIS, HYMAP, TIMS and MASTER (Abst). In *Proceedings, 14th Thematic Conference, Applied Geologic Remote Sensing*, November 6–8, 2000, Las Vegas, NV. Ann Arbor, MI, Environmental Research Institute of Michigan (ERIM), p. 122.
- Kruse, F. A., 2002. Combined SWIR and LWIR mineral mapping using MASTER/ASTER. In *Proceedings, IGARSS 2002*, June 24–28, 2002, Toronto, Canada. (Published on CD ROM – Paper Int1_B15_04, ISBN: 0-7803-7537-8. Also in hardcopy, v. IV, p. 2267–2269, IEEE Operations Center, Piscataway, NJ).
- Kruse, F. A., 2004. Comparison of ATREM, ACORN, and FLAASH atmospheric corrections using low-altitude AVIRIS data of Boulder, Colorado. In *Proceedings 13th JPL Airborne Geoscience Workshop, Jet Propulsion Laboratory*, 31 March–2 April 2004, Pasadena, CA.
- Kruse, F. A., 2012. Mapping surface mineralogy using imaging spectrometry. *Geomorphology*, **137**(1), 41–56.
- Kruse, F. A., Lefkoff, A. B., and Dietz, J. B., 1993. Expert system-based mineral mapping in northern Death Valley, California/Nevada using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, **44**, 309–336. Special issue on AVIRIS, May–June 1993.
- Kruse, F. A., Boardman, J. W., and Huntington, J. F., 1999. Fifteen Years of hyperspectral data: northern Grapevine Mountains, Nevada. In *Proceedings of the 8th JPL Airborne Earth Science Workshop: Jet Propulsion Laboratory Publication, JPL Publication 99–17*, pp. 247–258.
- Kruse, F. A., Boardman, J. W., and Huntington, J. F., 2003. Evaluation and validation of EO-1 hyperion for mineral mapping. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6), 1388–1400, doi:10.1109/TGRS.2003.812908. <http://dx.doi.org/10.1109/TGRS.2003.812908> (September 2012).
- Kruse, F. A., Perry, S. L., and Caballero, A., 2006. District-level mineral survey using airborne hyperspectral data, Los Menucos, Argentina. *Annals of Geophysics (Annali di Geofisica)*, **49**(1), 83–92.
- Lang, H. R., and Nadeau, P. H., 1984. Petroleum commodity report. In Abrams et al. (eds.), *The Joint NASA/GeoSat Test Case Project Final Report*. Tulsa, OK: The American Association of Petroleum Geologists Special Publication 2, part 2, pp. 10–1 to 10–28.
- Lang, H. R., Alderman, W. H., and Sabins, F. F., 1984. Patrick Draw, Wyoming, Petroleum test case report. In Abrams (ed.), *The Joint NASA/GeoSat Test Case Project Final Report*. Tulsa, OK: The American Association of Petroleum Geologists Special Publication 2, part 2, pp. 11–1 to 11–112.
- Lang, H. R., Adams, S. L., Conel, J. E., McGuffie, B. A., Paylor, E. D., and Walker, R. E., 1987. Multispectral remote sensing as stratigraphic tool, Wind River Basin and Big Horn Basin areas, Wyoming. *American Association of Petroleum Geologists Bulletin*, **71**(4), 389–402.
- Lyon, R. J. P., 1964. Evaluation of infrared spectrophotometry for compositional analysis of lunar and planetary soils, Part II: Rough and Powdered Surfaces. In *NASA contractor Report CR-100*, 262 p. Available at http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19650001173_1965001173.pdf (September 2012).
- MacDonald, I. R., Guinasso, N. L., Ackleson, S. G., Amos, J. F., Duckworth, R., Sassen, R., and Brooks, J. M., 1993. Natural oil slicks in the Gulf of Mexico visible from space. *Journal of Geophysical Research*, **98**(C9), 16351–16364.
- MacDonald, I. R., Reilly, J. F., Best, S. E., Venkataramaiah, R., Sassen, R., Guinasso, N. L., and Amos, J., 1996. Remote sensing inventory of active seeps and chemosynthetic communities in

- the Northern Gulf of Mexico. Hydrocarbon migration and its near-surface expression. *AAPG Memoir*, **66**, 27–37.
- Mars, J. C., and Rowan, L. C., 2006. Regional mapping of phyllic- and argillic-altered rocks in the Zagros magmatic arc, Iran, using Advanced Spaceborne Thermal Emission and Reflection Radiometers (ASTER) data and logical operator algorithms. *Geosphere*, **2**(3), 161–186, doi:10.1130/GES00044.1.
- McCubbin, D. G., and Brady, M. J., 1963. Depositional environment of the Almond reservoirs, Patrick Draw field, Wyoming. *The Mountain Geologist*, **6**, 3–36.
- NASA Goddard EO-1, 2012. Website: <http://eo1.gsfc.nasa.gov/> (September, 2012).
- Nevada Bureau of Mines (2008). (<http://www.nbmng.unr.edu/geo-thermal/site.php?sid=steamboathotsprings>) (September 2012).
- Pearlman, J. S., Barry, P. S., Segal, C. C., Shepanski, J., Beiso, D., and Carman, S. L., 2003. Hyperion, a space-based imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6), 1160–1173, doi:10.1109/TGRS.2003.815018. <http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=1220223> (September 2012) (September 2012).
- Pieters, C. M., and Mustard, J. M., 1988. Exploration of crustal/mantle material for the earth and moon using reflectance spectroscopy. *Remote Sensing of Environment*, **24**, 151–178.
- Prost, G. L., 2001. *Remote Sensing for Geologists*, 2nd edn. New York: Gordon and Breach, p. 374. ISBN 90-5702-629-5.
- Reston, M., and Cocks, T., 1998. Mapping mineralogy of the Mt. Fitton area, Flinders Ranges, South Australia, using HyMap airborne imaging spectrometer data. In *Proceedings 9th Australasian Remote Sensing Conference*, Sydney, July 1998.
- Richers, D. M., Reed, R. J., Horstman, K. C., Michels, G. D., Baker, R. N., Lundell, L., and Marrs, R. W., 1982. Landsat and soil-gas geochemical study of Patrick Draw Oil Field, Sweetwater County, Wyoming. *American Association of Petroleum Geologists Bulletin*, **66**, 903–922.
- Richers, D. M., Jones, V. T., Matthews, M. D., Maciolek, J., Pirkle, R. J., and Sides, W. C., 1986. The 1983 Landsat soil-gas geochemical survey of Patrick Draw Area, Sweetwater County, Wyoming. *American Association of Petroleum Geologists Bulletin*, **70**, 869–887.
- Rowan, L. C., 1998. Analysis of simulated advanced spaceborne thermal emission and reflection (ASTER) radiometer data of the Iron Hill, Colorado, study area for mapping lithologies. *Journal of Geophysical Research*, **103**(D24), 291–232.
- Rowan, L. C., and Mars, J. C., 2003. Lithologic mapping in the mountain pass, California area using advanced spaceborne thermal emission and reflection spectrometer (ASTER) data. *Remote Sensing of Environment*, **84**, 350–366, doi:10.1016/S0034-4257(02)00127-X. [http://dx.doi.org/10.1016/S0034-4257\(02\)00127-X](http://dx.doi.org/10.1016/S0034-4257(02)00127-X) (September 2012).
- Rowan, L. C., Bowers, T. L., Crowley, J. K., Anton-Pacheco, C., Gumiel, P., and Kingston, M. J., 1996. Analysis of airborne visible-infrared imaging spectrometer (AVIRIS) data of the Iron Hill, Colorado, carbonatite-alkalic igneous complex. *Economic Geology and the Bulletin of the Society of Economic Geologists*, **90**, 1966–1982.
- Rowan, L. C., Hook, S. J., Abrams, M. J., and Mars, J. C., 2003. Mapping hydrothermally altered rocks at Cuprite, Nevada, using the advanced spaceborne thermal emission and reflection radiometer (ASTER), a new satellite-imaging system. *Economic Geology and the Bulletin of the Society of Economic Geologists*, **98**(5), 1019–1027.
- Rowan, L. C., Simpson, C. J., and Mars, J. C., 2004. Hyperspectral analysis of the ultramafic complex and adjacent lithologies at Mordo NT, Australia. *Remote Sensing of Environment*, **91**, 419–431.
- Rowan, L. C., Simpson, C. J., and Mars, J. C., 2005. Lithologic mapping of the Mordor, NT, Australia ultramafic complex by using the advanced spaceborne thermal emission and reflection radiometer (ASTER). *Remote Sensing of Environment*, **99**, 105–126.
- Sabins, F. F., 1997. *Remote Sensing Principles and Interpretation*, 3rd edn. Long Grove: Waveland, p. 494. ISBN 1577665074.
- Salisbury, J. W., and D'Aria, D. M., 1992. Emissivity of terrestrial materials in the 8–14 μm atmospheric window. *Remote Sensing of Environment*, **42**, 83–106.
- Salisbury, J. W., Walter, L. S., Vergo, N., and D'Aria, D. M., 1988. *Infrared (2.1–13.5 micrometers) Spectra*. Washington, DC: U.S. Govt. Printing Office. U. S. Geological Survey Open-file Report 88–686.
- Salisbury, J. W., Walter, L. S., Vergo, N., and D'Aria, D. M., 1992. *Infrared (2.1–25 micrometers) Spectra of Minerals*. Baltimore, MD: Johns Hopkins University Press. 294 p.
- Schoen, R., White, D. E., 1967. *Hydrothermal Alteration of Basaltic Andesite and Other Rocks in Drill Hole GS-6, Steamboat Springs, Nevada*. Washington, DC: U.S. Govt. Printing Office. U.S. Geological Survey Professional Paper 575-B, pp. 110–119.
- Schoen, R., White, D. E., and Hemley, J. J., 1974. Argillization by descending acid at Steamboat Springs, Nevada. *Clays and Clay Minerals*, **22**, 1–22.
- Scott, L. F., McCoy, R. M., and Wulstein, L. H., 1989. Anomaly may not reflect hydrocarbon seepage: Patrick Draw field, Wyoming. *AAPG Bulletin*, **73**(7), 925–934. revisited: American Association of Petroleum Geologists.
- Sigvaldason, G. E., and White, D. E., 1962. *Hydrothermal Alteration in Drill Holes GS-5 and GS-7, Steamboat Springs, Nevada*. Washington, DC: U.S. Govt. Printing Office. U.S. Geological Survey Professional Paper 450-D, pp. D113–D117.
- Silberman, M. L., White, D. E., Keith, T. E. C., and Docktor, R. D., 1979. *Duration of hydrothermal activity at Steamboat Springs, Nevada, from ages of the spatially associated volcanic rock*. Washington, DC: U.S. Govt. Printing Office. U. S. Geological Survey Professional Paper 458-D. 14 p.
- Swayze, G. A., Smith, K. S., Clark, R. N., Sutley, S. J., Pearson, R. M., Vance, J. S., Hageman, P. L., Briggs, P. H., Meier, A. L., Singleton, M. J., and Roth, S., 2000. Using imaging spectroscopy to map acidic mine waste. *Environmental Science and Technology*, **34**, 47–54.
- Taranik, D. L., and Kruse, F. A., 1989. Iron mineral reflectance in geophysical and environmental research imaging spectrometer (GERIS) data. In *Proceedings, International Symposium on Remote Sensing of Environment*, Thematic Conference on Remote Sensing for Exploration Geology, 7th, October 2–6, 1989, Calgary, Alberta, Canada. Ann Arbor: Environmental Research Institute of Michigan, pp. 445–458.
- Ungar, S., Pearlman, J., Mendenhall, J., and Reuter, D., 2003. Overview of the Earth observing one (EO-1) mission. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(6), doi:10.1109/TGRS.2003.815999. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1220222> (September 2012).
- USGS EO-1, 2012. Website: <http://eo1.usgs.gov/> (September 2012).
- van der Meer, F., van Dijk, P., van der Werff, H., and Yang, H., 2002. Remote sensing and petroleum seepage; a review and case study. *Terra Nova*, **14**(1), 1–17.
- Weimer, R. J., 1966. Time-stratigraphic analysis and petroleum accumulations, Patrick Draw field, Sweetwater County, Wyoming. *AAPG Bulletin*, **50**(10), 2150–2175.
- White, D. E., 1955. Thermal springs and epithermal ore deposits. *Economic Geology*, **15**, 99–154.
- White, D. E., 1967. Some principles of geyser activity, mainly from Steamboat Springs, Nevada. *American Journal of Science*, **265**(8), 641–684.

- White, D. E., 1968. *Hydrology, Activity, and Heat Flow of the Steamboat Springs Thermal Systems, Washoe County, Nevada*. Washington, DC: U.S. Govt. Printing Office. U. S. Geological Survey Professional Paper 458-C. 109 p.
- White, D. E., 1981. Active geothermal systems and hydrothermal ore deposits. *Economic Geology*, **75**, 392–423.
- White, D. E., Anderson, E. T., and Grubbs, D. K., 1963. Geothermal brine well/mile-deep drill hole may tap ore-bearing magmatic water and rocks undergoing metamorphism. *Science*, **139**, 919–922.
- White, D. E., Thompson, G. A., and Sanberg, C. S., 1964. *Rocks, Structure, and Geologic History of Steamboat Springs thermal area, Washoe County, Nevada*. Washington, DC: U.S. Govt. Printing Office. U. S. Geological Survey Professional Paper 458-B. 63 p.
- White, D. E., Heropoulos, C., and Fournier, R. O., 1992. *Gold and other minor elements associated with the hot springs and geysers of Yellowstone National Park, Wyoming, supplemented with data from Steamboat Springs, Nevada*. Denver, CO: U.S. Geological Survey. U.S. Geological Survey Bulletin 2001. 19 p.
- Yamaguchi, A. B., Kahle, H., Tsu, T. K., and Pniel, M., 1998. Overview of advanced spaceborne thermal emission reflectance radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 1062–1071, doi:10.1109/36.700991. <http://dx.doi.org/10.1109/36.700991> (September 2012).

Cross-references

[Commercial Remote Sensing](#)
[Data Processing, SAR Sensors](#)
[Emerging Technologies](#)
[Geomorphology](#)
[Land Surface Emissivity](#)
[Observational Platforms, Aircraft, and UAVs](#)
[Observational Systems, Satellite](#)
[Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties](#)
[Reflected Solar Radiation Sensors, Multiangle Imaging](#)
[Remote Sensing, Physics and Techniques](#)
[Thermal Radiation Sensors \(Emitted\)](#)

S

SAR-BASED BATHYMETRY

Han Wensink¹ and Werner Alpers²

¹ARGOSS BV, Vollenhove, The Netherlands

²Institute of Oceanography, University of Hamburg, Hamburg, Germany

Synonyms

Retrieval of underwater bottom topography by using synthetic aperture radar (SAR)

Definition

Synthetic aperture radar (SAR) is an active instrument which yields high-resolution images in the microwave frequency band. When applied to image the ocean, it detects small variations in the small-scale sea surface roughness. These roughness variations can result, for example, from variations in surface current speed caused by a tidal flow over shallow sandbanks. Although the microwaves cannot penetrate into the water body, SAR can retrieve information on underwater bottom topography (or bathymetry) indirectly by measuring variations in the small-scale sea surface roughness.

Introduction

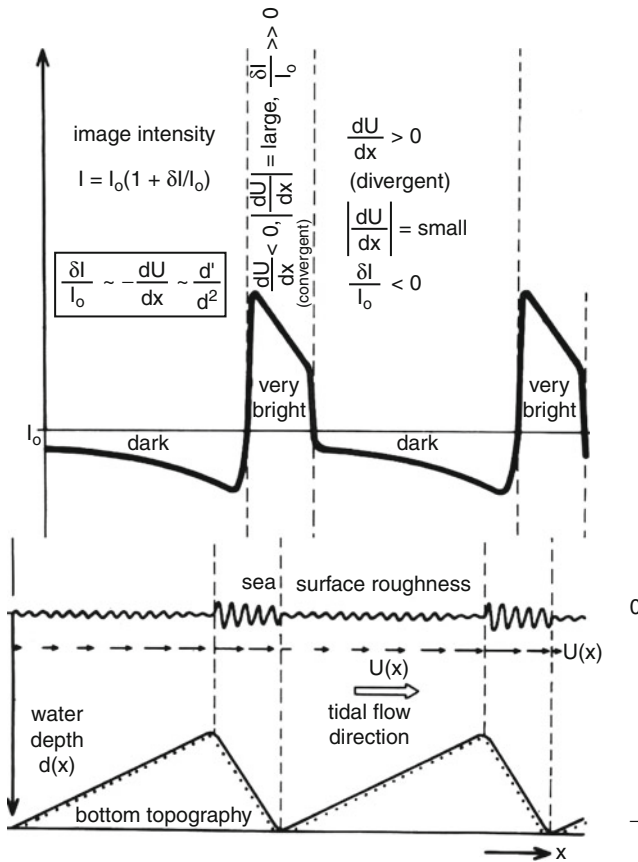
Underwater bottom topographic features, such as sandbanks and ridges, become visible on radar images of the sea surface only when there is a current present (usually a tidal current) which flows over them. This causes local perturbations to the current which in turn modulates the sea surface roughness. Locating underwater sandbanks by roughness variations of the sea surface has been used for several hundred years by mariners to avoid running aground. Since SAR is a very sensitive roughness sensor,

it is an ideal instrument for mapping the roughness pattern induced by (tidal) flow over underwater bottom topography.

However, it is not straightforward to invert SAR images of the sea surface into maps of underwater bottom topography. Although theories describing the radar imaging of underwater bottom topography have been developed (Alpers and Hennings, 1984; Shuchman et al., 1985; Romeiser and Alpers, 1997; Vogelzang, 1997), they are at present not capable of generating depth charts (maps of underwater bottom topography) from SAR data alone. But if they are combined with acoustic sounding (sonar) data, they are of great use in generating depth charts. This is achieved by using SAR images to interpolate between depth profiles measured by acoustic sounders operated from ships. Thus, radar measurements can serve to reduce greatly the number of ship tracks required for generating bathymetric maps of a given accuracy and thus reduce greatly the costs of bathymetric surveying.

Principle of radar imaging of underwater bottom topography

Imaging radars such as real aperture radar (RAR) or synthetic aperture radar (SAR) sense underwater topographic features indirectly by variations of the small-scale sea surface roughness induced by a variable surface current. The radar imaging mechanism of underwater bottom topography consists of three parts: (1) the modulation of the current by the structure of the underwater bottom topography, (2) the modulation of the sea surface waves by the variable surface current, and (3) the interaction of the microwaves with the surface waves. This is depicted schematically in [Figure 1](#).



SAR-Based Bathymetry, Figure 1 Schematic plot showing the relationship between radar image intensity, sea surface roughness, tidal flow, and underwater bottom topography (bathymetry). In this case, the bathymetry consists of an asymmetrically shaped sand wave. The steep slopes of the sand wave face the flow direction and are associated with a strongly reduced image intensity. (Plot reproduced from Alpers and Hennings, 1984).

The tidal current is modified by the variable water depth as indicated by the arrows. The current velocity at the sea surface is higher over shallow areas than over deep areas giving rise to convergent flow regimes, where the sea surface roughness is increased, and to divergent flow regimes, where roughness is decreased. In the convergent areas, the waves are “squeezed,” while in the divergent areas they are “stretched.” This leads, in the convergent areas, to an increase of the amplitude of the Bragg waves (which are responsible for the radar backscattering through a process of resonance between the wavelengths of short waves on the ocean surface and the obliquely incident radar beam) and thus to an increase of the backscattered radar power, and in the divergent areas it leads to a decrease. Therefore, on radar images, the convergent areas appear as areas of enhanced image intensity and the divergent areas as areas of reduced image



SAR-Based Bathymetry, Figure 2 SEASAT SAR image of the north eastern approach to the English Channel acquired August 19, 1978. Visible is on the right-hand side the French/Belgian coast. (Because this image is not corrected for the variation of brightness with incident angle, the image intensity is high along the bottom of the image corresponding to near range (steep incident angles)). The imaged area is 100 x 100 km.

intensity. In most cases, the surface current is a tidal current. The strongest modulation of the sea surface roughness, and thus of the backscattered power, is obtained when the tidal velocity is at its peak.

Examples of SAR images showing underwater topographic features

In this section, we present three examples of SAR images showing underwater topographic features acquired by the L-band (1.275 GHz) SAR aboard the American SEASAT satellite (launched 1978) and the C-band (5.3 GHz) SAR aboard the European ERS-1 satellite (launched 1991).

Figure 2 shows a SEASAT SAR image of the northeastern approach to the English Channel. On the right-hand side of the image, the French-Belgian coast is visible with the French town of Calais at the bottom. The V-shaped feature in the lower left-hand section of the image is a pair of two underwater ridges called South Falls (the thin line to the left) and Sandettie (the broader line to the right). South Falls is about 30 km long and 600–800 m wide and rises to within 7 m of the sea surface. The seafloor between the two ridges has a depth between 30 and 40 m. At the time of the SAR data acquisition, the tidal flow was directed toward the southwest with a speed between 1.7



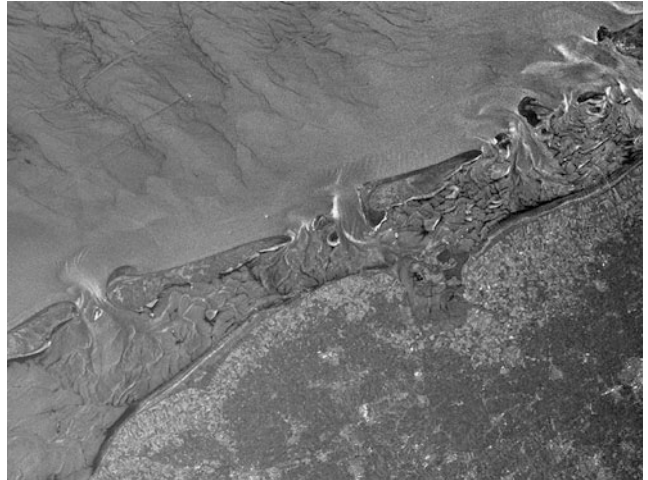
SAR-Based Bathymetry, Figure 3 ERS-1 SAR image acquired July 8, 1995, showing the Xinchuan Gang (Subei) Shoal off the coast of China (north of Shanghai). The dark areas off the coast are sand banks which have fallen dry during ebb tide. The imaged area is 100×100 km. ©ESA 1995.

and 2.4 ms^{-1} . The wind speed was 4 ms^{-1} . Thus, optimum SAR imaging conditions were encountered: the tidal current velocity was close to its peak value during the tidal cycle, and the wind speed was well above threshold for small-scale wave generation (above $2\text{--}3 \text{ ms}^{-1}$), but not too high (above 8 ms^{-1}) such that the sea surface roughness modulation caused by the variable current over the sandbanks would be masked by wind effects.

Figure 3 shows an ERS-1 SAR image of the nearshore section of the Xinchuan Gang Shoal off the coast of China (north of Shanghai). The light gray area on the left-hand side and on the bottom of the image is land. The irregular dark band adjacent to the coast and the dark areas farther off the coast are tidal flats that have fallen dry. The light gray patterns in the right-hand section of the image are submerged sand ridges, which are quite similar to the underwater sand ridges visible in Figure 2.

Figure 4 shows an ERS-1 SAR image of the Waddenze in the North Sea off the Dutch coast. At the time of the SAR data acquisition, the tidal current was close to its peak and the wind was blowing from north with a speed of about 5 ms^{-1} .

Visible are (from left to right) the Dutch islands Terschelling, Ameland, and Schiermonnikoog. The Plaatgat



SAR-Based Bathymetry, Figure 4 ERS-1 SAR image of the Waddenze acquired August 3, 1995, at 1035 UTC. Between the land (lower section of the image) and the open sea (upper section) several islands are located: The Plaatgat area is located approximately in the center of the image. The imaged area is 100×100 km. ©ESA 1995.

area, located north of the tidal inlet between the islands of Ameland and Schiermonnikoog, is a shallow sea area which requires frequent updates of bathymetric maps because here the bottom topography is highly variable. In this area SAR images have been used for generating of bathymetric maps.

Generation of bathymetric maps using SAR images

In certain areas (e.g., the Dutch and the German coasts) the location, shape, and depth of underwater banks or shoals are highly variable due to the presence of strong currents and to the presence of sand or gravel at the seafloor. Under these conditions, regular monitoring and dredging are required in coastal areas. Conventionally, depth charts are obtained by sonar measurements carried out from dedicated vessels (survey vessels). Since depth measurements carried out from ships are quite expensive, it is a challenge to find a way to reduce the number of ship tracks without reducing the accuracy of the depth measurements. Here, SAR imagery comes into play, as it can be used to interpolate depth values between tracks sailed by the survey vessel.

A method to incorporate SAR images in conventional bathymetric surveying has been developed by ARGOS in the Netherlands called the Bathymetry Assessment System (BAS; Calkoen et al., 2001). It constructs depth maps from radar images and a limited number of echo soundings by numerical inversion of a two-dimensional model of the imaging mechanism (Vogelzang et al., 1997; Wensink and Campbell, 1997). For inverting SAR

images into bathymetric maps, the following additional data are needed: (1) sonar sounding data along calibration tracks to allow tuning, (2) tidal data relating to the acquisition time of the SAR image, (3) wind speed and direction relating to the acquisition time of the SAR image, and (4) the SAR image. In the inversion scheme, the depth profiles are retrieved from the SAR data in an iterative process such that the closest possible correspondence between modeled and measured depth profiles is obtained. In several validation experiments, it has been shown that this inversion scheme yields fine-grid bathymetric maps with depth accuracy better than 30 cm.

Summary

Although microwaves cannot penetrate into the water body, SAR can yield information on shallow-water bathymetry or underwater bottom topography. This is achieved indirectly by means of sensing variations in the sea surface roughness over bathymetry when a strong current (usually tidal current) is present. SAR is a very sensitive instrument to sense small variations in the (small-scale) sea surface roughness image. At present, it is not possible to invert reliably SAR images, which are related to sea surface roughness maps, into bathymetric maps by using theoretical models. However, if SAR images are combined with acoustic sounding data acquired from ships, they are of great value for generating bathymetric maps and thus can help to reduce greatly the cost of updating depth charts in tidal areas.

Bibliography

- Alpers, W., and Hennings, I., 1984. A theory of the imaging mechanism of underwater bottom topography by real and synthetic aperture radar. *Journal of Geophysical Research*, **89**, 10529–10546.
- Calkoen, C. J., Hesselmanns, G. H. F. M., Wensink, G. J., and Vogelzang, J., 2001. The bathymetry assessment system: efficient depth mapping in shallow seas using radar images. *International Journal of Remote Sensing*, **22**, 2973–2998.
- <http://earth.esa.int/applications/ERS-SARtropical/oceanic/bottom/index.html>
- http://www.sarusersmanual.com/ManualPDF/NOAASARManual_CH10_pg245-262.pdf
- Romeiser, R., and Alpers, W., 1997. An improved composite surface model for the radar backscattering cross section of the ocean surface, 2. Model response to surface roughness variations and the radar imaging of underwater bottom topography. *Journal of Geophysical Research*, **102**, 25251–25267.
- Shuchman, R. A., Lyzenga, D. R., and Meadows, G. A., 1985. Synthetic aperture radar imaging of ocean-bottom topography via tidal-current interactions: theory and observations. *International Journal of Remote Sensing*, **6**, 1179–1200.
- Vogelzang, J., 1997. Mapping submarine sand waves with multi-band imaging radar, 1. Model development and sensitivity analysis. *Journal of Geophysical Research*, **102**, 1163–1181.
- Wensink, H., and Campbell, G., 1997. Bathymetric map production using the ERS SAR. *Backscatter*, **8**, 17–22.

Cross-references

[Microwave Surface Scattering and Emission Observational Systems, Satellite Radar, Synthetic Aperture Radars](#)

SEA ICE ALBEDO

Donald Perovich
USACE Cold Regions Research and Engineering Laboratory, Hanover, NH, USA

Synonyms

Frozen oceans; Sunlight reflection

Definition

Albedo – the fraction of sunlight reflected by a surface.
Bidirectional reflectance distribution function – the angular distribution of light reflected by a surface as a function of sun angle.

Irradiance – sunlight integrated over a hemisphere with units of watts per meter squared.

Leads – areas of open ocean in a sea ice cover.

First year ice – ice that has not yet endured an entire summer melt season.

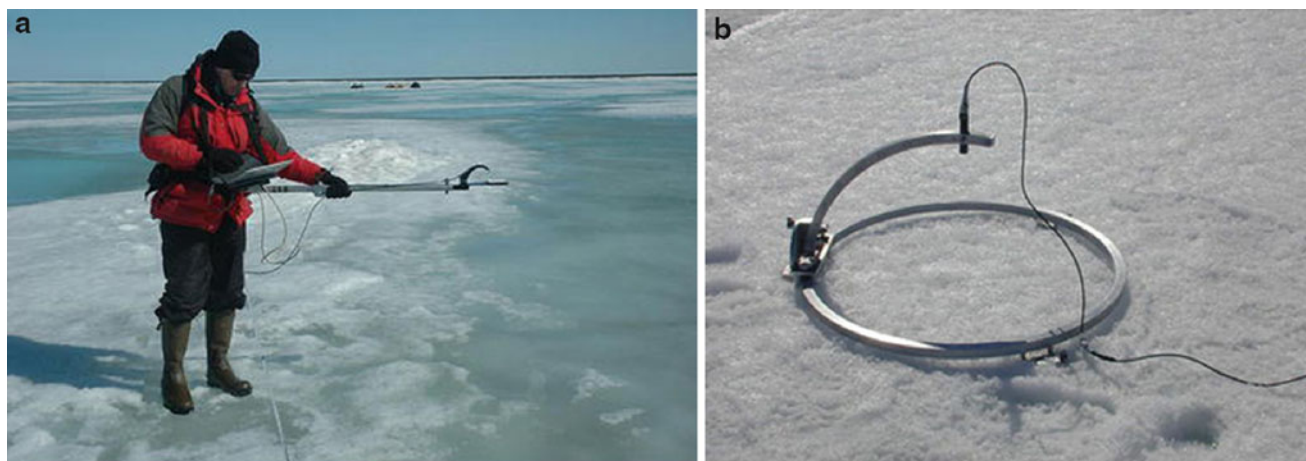
Multiyear ice – ice that is one or more years old.

Introduction

The reflection of sunlight from sea ice is of considerable importance for the interpretation of remote sensing imagery and for the large-scale surface heat budget of the ice cover. Light reflection can be characterized by two terms: the albedo and the bidirectional reflectance distribution function (BRDF). These quantities can either be spectral or integrated over the solar spectrum (wavelengths from 280 to 2,500 nm).

The albedo is simply the fraction of the incident solar irradiance that is reflected by the surface. It is dimensionless. The irradiance is integrated over a hemisphere (downwelling for incident and upwelling for reflected) and has units of watts per meter squared. A surface that is a perfect absorber has an albedo of 0, while a perfect reflector has an albedo of 1. The albedo is a key parameter in the surface heat budget of sea ice and has climate implications (Curry et al., 1993). [Figure 1a](#) illustrates using a field portable spectroradiometer to measure spectral albedo. When measuring albedo, it is important to select a shadow-free, undisturbed surface and to keep the detector level while recording the incident and reflected irradiance.

The BRDF is the angular distribution of radiance reflected from a surface as a function of solar incident angle. It is often used for interpreting remote sensing imagery. Satellite sensors measure reflectance at



Sea Ice Albedo, Figure 1 (a) Measuring albedo using a field portable spectroradiometer. As oriented, the instrument is measuring the incident irradiance. The surface is a recently formed melt pond; (b) Apparatus for measuring BRDF consisting of two rings that can be rotated to any combination of zenith and azimuth angle. The surface is bare white ice.

a particular angle for a particular solar incident angle. The reflectances are then interpreted to generate geophysical quantities such as albedo and the type of ice present. An apparatus for measuring BRDF is presented in Figure 1b. It consists of horizontal and vertical rings that can be oriented to any azimuth and zenith angle. A probe is mounted in the ring as illustrated. Surface reflectances are measured at many angles and normalized to a white reference standard.

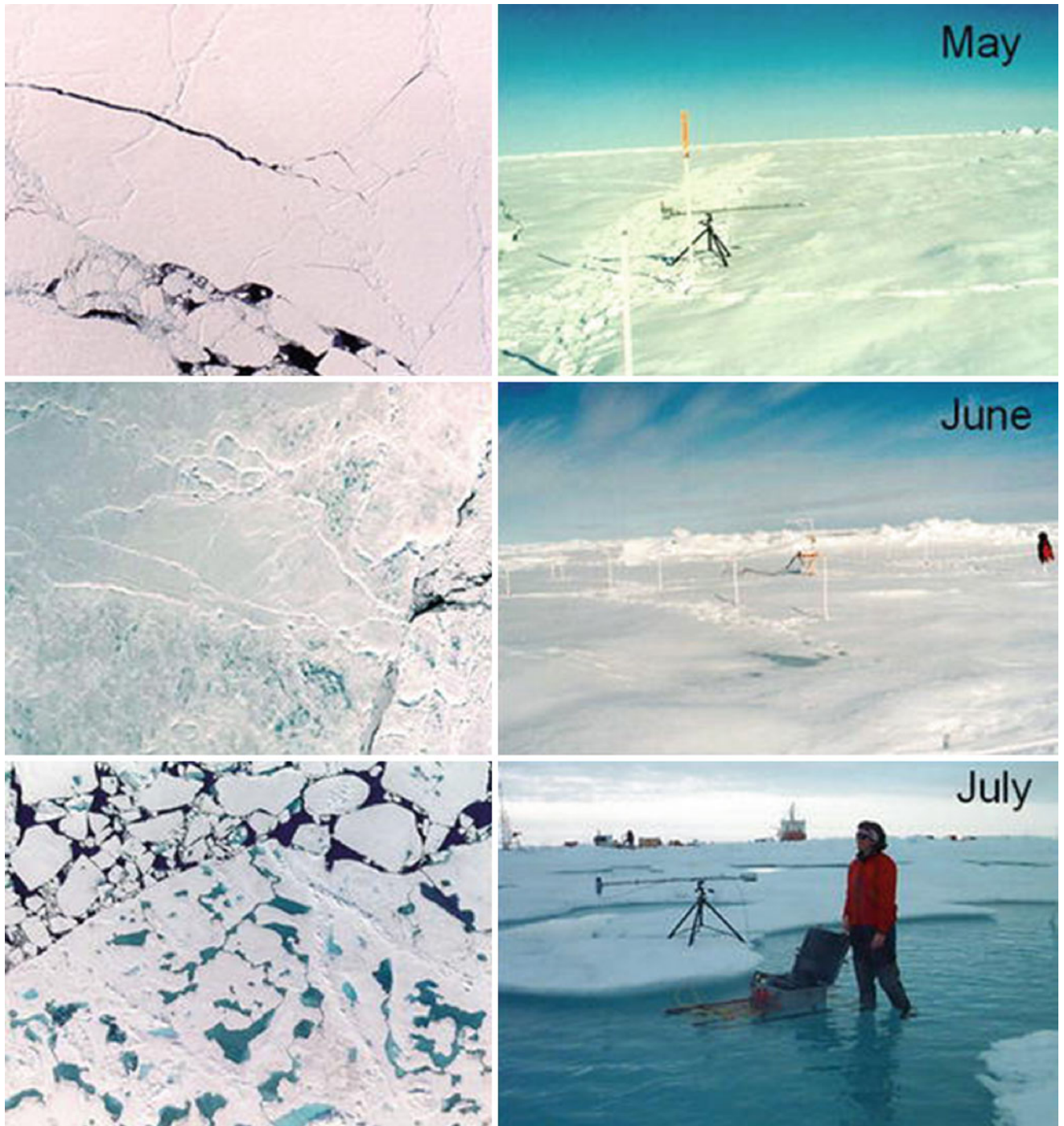
Impact of surface conditions

Light reflection from sea ice is governed largely by surface conditions (Perovich, 1996). As Figure 2 demonstrates, sea ice surface conditions vary spatially and temporally. Before the summer melt season begins, there are two primary surface types present: snow-covered ice and leads (areas of open ocean). As summer progresses, the surface becomes a variegated mixture of melting snow, bare ice, melt ponds, and leads. The blue areas in the photographs are melt ponds, where melt water collects in a surface layer a few centimeters to half a meter deep. The melt ponds evolve throughout the summer growing deeper and developing a complex, interconnected drainage system. In August or September, fall freeze-up begins and the surface consists of snow-covered ice, young, thin ice, and leads.

Because of the importance of albedo to the surface heat budget, albedo measurements have long been a routine part of sea ice field experiments. Photographs of eight distinct surface types, along with their wavelength-integrated and spectral albedos are presented in Figure 3 (Allison et al., 1993; Grenfell and Maykut, 1977). The varied surface conditions result in albedos that range from a minimum of 0.07 for leads to a maximum of 0.85 for

cold snow-covered ice. Sea ice albedos cover almost the entire possible range of 0–1. The differences in albedo are due to scattering. For example, snow, with an abundance of air-ice interfaces, is a highly scattering medium, while clear melt water scatters little light. For most of the year, the predominant surface type is cold snow-covered ice. In late spring, the snow begins to melt and its albedo decreases to about 0.70–0.75. Once the snow has melted, bare ice is exposed. This bare ice typically has a few centimeter-thick surface layer consisting of small fragments of deteriorated ice. Ice that has been through a full annual cycle (multiyear ice) typically has more air bubbles and more scattering than ice that has not (first year ice). Bare multiyear ice has an albedo of $0.65 (\pm 0.05)$ compared to a value of $0.55 (\pm 0.05)$ for first year ice. Melt pond albedos range from 0.1 to 0.4.

Spectral albedos for eight surface types are presented in Figure 3b (Perovich et al., 2002; Winther et al., 2004). The spectral differences in albedo are due to absorption. Light absorption in ice and water increases rapidly with wavelength, so there is a general tendency for albedo to decrease as wavelength increases. The albedos of snow and bare ice are large and vary little with wavelength across the visible spectrum (400–750 nm) and these ice types appear white. Leads have a small albedo across the visible spectrum and look black. Most melt ponds have a sharp spectral peak and are shades of blue, azure, and cobalt. The brightness and color of the melt ponds are largely determined by the properties of the underlying ice. Dark ponds typically have a thinner layer of underlying ice, while whitish ponds have thicker ice with more air bubbles. Blue ponds have relatively thick ice (more than 1 m), with few air bubbles. The color is a result of minimal absorption of light in ice at blue wavelengths.



Sea Ice Albedo, Figure 2 (Continued)



Sea Ice Albedo, Figure 2 Aerial and surface-based photographs of sea conditions from May through September.

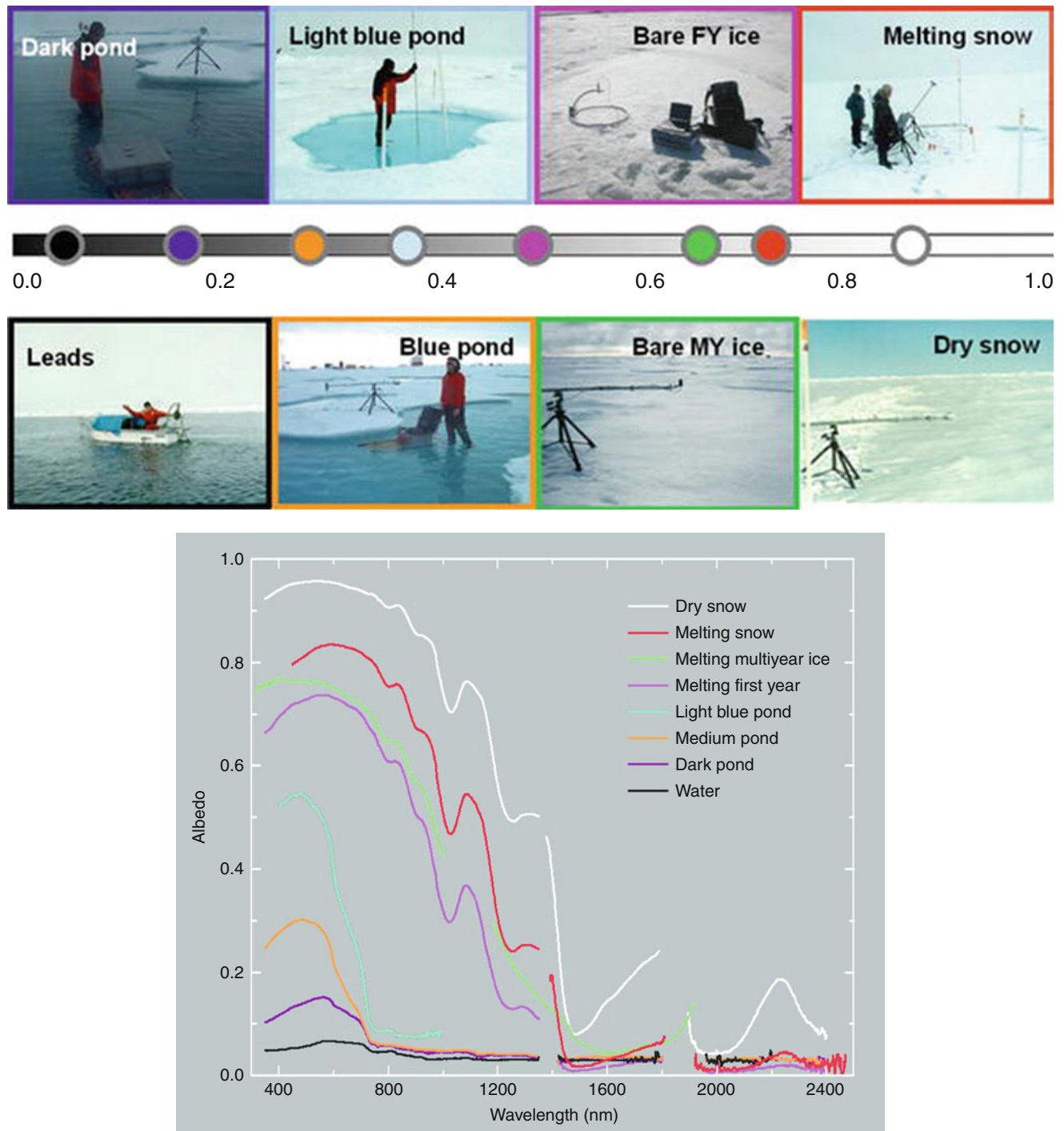
Ice albedo feedback

The importance of sea ice albedo transcends the polar regions and has implications for the global climate system through the ice albedo feedback. There is an extreme contrast in albedo between snow-covered ice and leads. Also, as the photographs in [Figure 2](#) demonstrate, there is a general decrease in albedo as the sea ice melts. These phenomena combine to form the ice albedo feedback. In spring, the surface is primarily snow-covered ice, with a small area of leads. The albedo is large and roughly 80 % of the sunlight is reflected. However, the remainder is absorbed and, with warming air temperatures, eventually results in surface melting. This melting lowers the albedo, which causes more melting and a further lowering

of albedo. This positive ice albedo feedback mechanism can amplify changes to the sea ice cover and consequently is important to climate change.

Conclusion

A marked decline in the amount of Arctic sea ice in summer has been observed in recent years at a rate faster than predicted by models (Stroeve et al., [2007](#), [2008](#)). There is evidence that there is enhanced solar heating of the upper ocean and consequently an increase in ice melting (Perovich et al., [2008](#); Steele et al., [2008](#)). The ice albedo feedback is contributing to the Arctic sea ice loss and may be accelerating it.



Sea Ice Albedo, Figure 3 Photographs of eight common sea ice surface types (top). Dots on the line represent the wavelength integrated albedo for these surface types, with the dot color matching the photograph border color. Spectral albedos for the eight surface types are shown below.

Bibliography

- Allison, I., Brandt, R. E., and Warren, S. G., 1993. East Antarctic sea ice: albedo, thickness distribution, and snow cover. *Journal of Geophysical Research*, **98**, 12417–12429.
- Curry, J. A., Schramm, J. L., and Ebert, E. E., 1993. Sea ice-albedo climate feedback mechanism. *Journal of Climate*, **8**, 240–247.
- Grenfell, T. C., and Maykut, G. A., 1977. The optical properties of ice and snow in the Arctic Basin. *Journal of Glaciology*, **18**, 445–63.
- Perovich, D. K., 1998. The optical properties of sea ice. In *Physics of Ice Covered Seas*. Helsinki: University of Helsinki Press, Vol. 1, p. 446.
- Perovich, D. K., Grenfell, T. C., Light, B., and Hobbs, P. V., 2002. The seasonal evolution of Arctic sea ice albedo. *Journal of Geophysical Research*, doi:10.1029/2000JC000438.
- Perovich, D. K., Richter-Menge, J. A., and Jones, K. F., 2008. Sunlight, water, and ice: extreme Arctic sea ice melt during the summer of 2007. *Geophysical Research Letters*, **35**, L11501, doi:10.1029/2008GL034007.
- Steele, M., Ermold, W., and Zhang, J., 2008. Arctic Ocean surface warming trends over the past 100 years. *Geophysical Research Letters*, **35**, L02614, doi:10.1029/2007GL031651.
- Stroeve, J., Holland, M. M. W., Meier, T. S., and Serreze, M., 2007. Arctic sea ice decline: faster than forecast. *Geophysical Research Letters*, **34**, L09501, doi:10.1029/2007GL029703.
- Stroeve, J., et al., 2008. Arctic sea ice extent plummets in 2007. *Eos, Transactions, American Geophysical Union*, **89**, 13.
- Winther, J. G., Edvardsen, K., Gerland, S., and Hamre, B., 2004. Surface reflectance of sea ice and under-ice irradiance in Kongsfjorden, Svalbard. *Polar Research*, **23**(1), 115–118, doi:10.1111/j.1751-8369.2004.tb00134.x.

Cross-references

[Cryosphere and Polar Region Observing System](#)
[Cryosphere, Climate Change Feedbacks](#)
[Cryosphere, Measurements and Applications](#)
[Sea Ice Concentration and Extent](#)
[Surface Radiative Fluxes](#)

SEA ICE CONCENTRATION AND EXTENT

Josefino C. Comiso

Cryospheric Sciences Laboratory, Code 615, Earth Sciences Division, NASA Goddard Space Flight Center, Greenbelt, MD, USA

Synonyms

Frozen sea water; Ice floe; Ice pack

Definitions

Sea ice. Sea ice is any form of ice which originated from the freezing of seawater and is found in the sea.

Sea ice concentration. Sea ice concentration is the fraction of an observational area covered completely by sea ice.

Sea ice extent. Sea ice extent represents the total area covered partially or totally by sea ice. For remote sensing applications, a lower threshold of 15 % in the fraction of sea ice cover is usually applied.

Sea ice

Among the most seasonal and most dynamic parameters on the surface of the Earth is sea ice which at any one time covers about 3–6 % of the planet. In the Northern Hemisphere, sea ice grows in extent from about $6 \times 10^6 \text{ km}^2$ to $16 \times 10^6 \text{ km}^2$, while in the Southern Hemisphere, it grows from about $3 \times 10^6 \text{ km}^2$ to about $19 \times 10^6 \text{ km}^2$ (Comiso, 2010; Gloersen et al., 1992). Sea ice is up to about 2–3 m thick in the Northern Hemisphere and about 1 m thick in the Southern Hemisphere (Wadhams, 2002), and compared to the average ocean depth of about 3 km, it is a relatively thin, fragile sheet that can break due to waves and winds or melt due to upwelling of warm water. Being constantly advected by winds, waves, and currents, sea ice is very dynamic and usually follows the directions of the many gyres in the polar regions. Despite its vast expanse, the sea ice cover was previously left largely unstudied and it was only in recent years that we have understood its true impact and significance as related to the Earth's climate, the oceans, and marine life.

We now know that sea ice is part of the polar heat sink and an important component of the climate system. It is a good insulator and serves to keep the ocean warm during winter and at the same time is a good reflector of solar radiation, thereby keeping the ocean relatively cold in the summer. Sea ice also redistributes salt in the ocean since salinity is enhanced where it forms and is reduced where it melts. Among the key sources of cold and high salinity water are regions where ice is formed at a relatively high rate, as in coastal polynyas. These polynyas are known to be ice factories and are the key sources of bottom water that helps maintain the global thermohaline circulation (Gordon and Comiso, 1988; Martin et al., 2007). In the deep ocean regions, the formation of sea ice can initiate deep ocean convection that enables vertical exchanges of chemicals and the resupply of surface water with nutrients and oxygen. Sea ice is also an integral part of the ecology of the polar environment. It is a habitat and a source of food for many marine organisms, and during the melt season, it causes the formation of low-density surface water layer that is stable and exposed to abundant sunlight. Such layers are ideal platforms for photosynthesis and have been regarded as potential sites of phytoplankton blooms (Smith and Comiso, 2008).

Modeling studies have shown that global warming signals can be amplified by as much as 3–5 times in the Arctic region (Holland and Bitz, 2003). This is mainly because of ice-albedo feedback effects associated with the high albedo of ice compared to that of open water. The process may already be occurring in a significant manner since the total Arctic sea ice cover has been declining at about 3–4 % per decade (Parkinson et al., 1999; Bjorgo et al., 1997), while the perennial ice (which is ice that survives the summer melt) has been declining at a much higher rate of about 9–13 % per decade (Comiso, 2002; Stroeve et al., 2004; Comiso et al., 2008), with

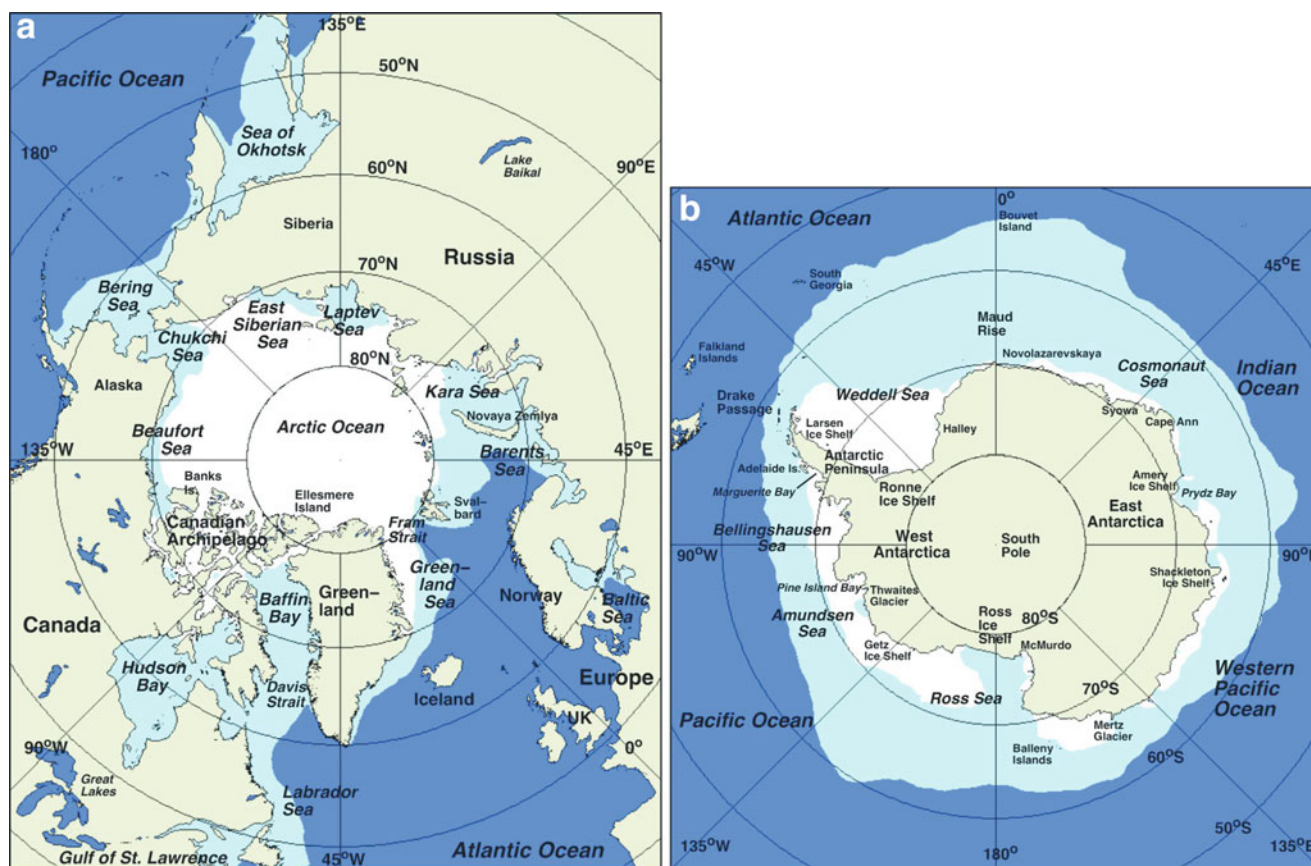
sudden declines in the summers of 2007 and 2012. On the other hand, the total extent of sea ice in the Antarctic is going the opposite way with the ice cover increasing at about 1 % per decade (Cavalieri et al., 1997; Zwally et al., 2002). The phenomenon in the Antarctic is still poorly understood and has been associated with a freshening of the Ross Sea that causes enhanced vertical stratification that in turn reduces heat fluxes originating from warmer water in the deeper layers (Zhang, 2007). More recent studies suggest that the depletion of stratospheric ozone in the Antarctic resulted in the deepening of lows in the West Antarctic region, causing strong southerly winds and, therefore, enhanced sea ice production off the Ross Sea ice shelf region (Turner et al., 2009).

Physical and radiative characteristics of sea ice

Geographical maps of the polar regions and the general location of the sea ice cover during minimum and maximum extents as derived from satellite data are shown in Figure 1 for both hemispheres. It is apparent that the

distributions of the sea ice cover in the two regions are different and are affected by the location of continental land masses and other environmental factors. In the Northern Hemisphere, land surrounds the Arctic Ocean which is covered almost completely by sea ice in winter. Sea ice is also found in peripheral seas and bays like the Okhotsk Sea, Bering Sea, Hudson Bay, Labrador Sea, and Baffin Bay during winter but is confined as an almost continuous sheet in the Arctic Basin in the summer. In the Southern Hemisphere, on the other hand, sea ice surrounds the continental land mass, called Antarctica, and has no northern boundary. The ice cover is also almost a continuous ice sheet in winter and gets fragmented at the end of the summer.

Supercooling of ocean water leads to the formation of tiny crystals of ice, called frazil ice, that accumulates at the surface. These tiny particles coagulate to form a soupy layer at the surface that looks like oil slick and has been referred to as grease ice. In the inner pack where the ocean surface is relatively stable, the ice layer gets thicker through the growth of congelation ice at the underside and forms an ice sheet, called nilas, which initially



Sea Ice Concentration and Extent, Figure 1 Location maps of the polar regions in (a) Northern and (b) Southern Hemispheres. Also shown are averages of ice concentrations during ice minimum (white) and ice maximum (light blue) for the period November 1978 to December 2009.

appears dark and then gets lighter gray with time as it gets thicker. Near the ice edge, where the surface is constantly perturbed by winds and waves, the grease ice gets transformed into loosely aggregated disks called pancake ice that have raised edges because of constant collisions with each other. The ice grows in thickness usually through thermodynamic growth, through rafting, which is a process in which one piece of ice goes on top of the other or through ridging which is when two thick ice floes collide and the thinner ice in between is broken up. When ice sheets become about 15–30 cm thick, they are called young ice and beyond that, they become first year ice which is the most dominant sea ice type during the winter period. The sea ice cover that survives one summer melt period is referred to as second year ice, and if it survives more than one summer, it is called multiyear ice.

The key parameter associated with the radiative characteristics of a material is its emissivity which for sea ice depends on many factors such as its thickness, salinity, snow cover, temperature, and liquid content. At microwave frequencies and according to the Rayleigh-Jeans approximation, the radiance from a surface is directly proportional to the temperature of the emitting material, with the emissivity being the proportionality constant. The high contrast in the microwave between the high emissivity of sea ice and low emissivity liquid water makes passive microwave sensors an effective tool for monitoring the sea ice cover. The contrast is most pronounced when the surface of the sea ice cover is dry and is dominated by young ice and thicker ice types. The emissivity of the new ice types changes depending on thickness and varies from that slightly higher than liquid water (as with grease ice) through intermediate values (as with gray ice) before it becomes young ice or thicker ice types. The emissivity of melting snow cover and meltponded areas is also unpredictable and causes ambiguities in interpretation. There is also a difference in the emissivity of seasonal and the older ice types. Young and first year sea ice types, which are the main components of seasonal ice, have relatively high salinity at up to about 12 psu (Weeks and Ackley, 1986). The salinity depends on the amount of brine that is entrapped within the ice during the process of ice-crystallization. Gradual desalinization follows mainly through a process called brine drainage that is usually accelerated during the summer (Martin, 2004; Tucker et al., 1992). Thus, the multiyear ice that survives two or more summers would have much lower salinities and hence different dielectric properties than those of seasonal ice. This phenomenon and volume scattering within the ice makes the multiyear ice emissivity distinctly lower than that of seasonal ice (Vant et al., 1974; Grenfell, 1992) and an opportunity to also monitor separately the multiyear ice cover. The dry snow cover is mainly transparent, but at some microwave frequencies, it also serves as a scatterer of radiation and can affect the emissivity, especially when the grain size becomes comparable to the wavelength of the radiation. The variations in emissivity for different ice or surface types can be a problem but

are mainly accounted for in the estimates of ice concentration as discussed in the following section.

Sea ice concentration: algorithms and validation

The advent of satellite imaging passive microwave sensors enabled for the first time quantitative assessments of the large-scale distribution and estimates of the true extent and area of the global sea ice cover (Zwally et al., 1983). This capability started with the Nimbus-5 Electrically Scanning Microwave Radiometer (ESMR), cross-scanning and horizontally polarized sensor, which operated at 19 GHz and was launched in December 1972. This was followed by the Scanning Multichannel Microwave Radiometer (SMMR), launched onboard the Nimbus-7 satellite in 1978 and subsequently by the series of DMSP Special Scanning Microwave Imager (SSM/I), the first of which was launched in July 1987. The most recent and likely the most versatile is the Advanced Microwave Scanning Radiometer which was launched on board EOS Aqua on May 4, 2002, (AMSR-E) and on May 18, 2012, (AMSR-2) on board JAXA GCOM-W. A comparison of the resolution and other technical characteristics of these sensors is provided by Lubin and Massom (2006).

Passive microwave sensors have the advantage over other types of sensors of providing day/night, almost all weather data globally at a good temporal resolution. Spatial resolution has been cited as a limitation of the system with historical data being archived at a resolution of 25 km. However, this limitation is minimized with the use of a mixing algorithm that estimates the fraction of sea ice within each measurement. The resolution has improved considerably with AMSR-E which provides standard ice data at a resolution of about 12 km, but data with a resolution of 6 km can be obtained from the dual-polarized 89 GHz channels. Other types of sensors have also been used for the study of sea ice such as Landsat, which provides visible imagery at 15 m resolution, and the Synthetic Aperture Radar (SAR) which provides active microwave images at a resolution of about 30 m. Data from these sensors have been useful for mesoscale studies and for validation of passive microwave data (Comiso et al., 2001). Global coverage from these high-resolution sensors, however, has been limited spatially and temporally and cannot be used effectively for large-scale variability studies of global sea ice cover. Several algorithms have been developed for SMMR and SSM/I data (Svendsen et al., 1983; Swift et al., 1985; Cavalieri et al., 1984; Comiso, 1986; Steffen et al., 1992). The standard algorithms used for estimating ice concentrations using AMSR-E data are the Bootstrap and NT2 Algorithms as described by Comiso et al. (2003). Despite differences in the techniques and channels utilized, the Bootstrap and NT2 Algorithms provide very similar values in ice concentration, ice extent, and ice area (Parkinson and Comiso, 2008; Comiso and Parkinson, 2008; Comiso and Nishio, 2008). In this section, we will

briefly describe how ice concentration is derived from passive microwave data.

The brightness temperature (T_B), recorded by a satellite sensor at a given wavelength, can be estimated using the basic radiative transfer equation given by

$$T_B = \varepsilon T_S e^{-\tau} + \int_0^{\tau} T(z) \zeta(z) e^{-\tau+\tau'(z)} d\tau'(z) + (1 - \varepsilon) \kappa e^{-\tau} - \int_0^{\tau} T(z) \zeta(z) e^{-\tau'(z)} d\tau'(z) \quad (1)$$

where ε is the emissivity of the surface, T_S is the physical temperature of the surface, $\tau'(z)$ and τ are the atmospheric opacities from the surface to a height z and from the surface to the satellite height (usually about 800 km), respectively, $T(z)$ is atmospheric temperature at a height z , κ is an estimate of the diffusiveness of the surface reflection, and $\zeta(z)$ is the emittance at z . In Equation 1, the first term in the right hand side of the equation represents radiation directly from the earth's surface, which is the dominant contribution for measurements at microwave frequencies. The second term represents satellite-observed radiation that emanates directly from the atmosphere, while the third term represents downwelling radiation from the atmosphere that has been reflected toward the satellite from the earth's surface. A fourth term that takes into account the reflected contribution of radiation from free space (i.e., the cosmological 2 K contribution from the Big Bang), which is an additive contribution, is negligible and not included in Equation 1.

The estimation of surface parameters from satellite data usually requires the effective implementation of a radiative transfer model that makes use of Equation 1. The goal is to correct for atmospheric effects and to obtain a good estimate of the brightness temperature (T_{BS}) of the surface of interest which in our case is the first term in Equation 1 without the exponential term. Getting an estimate of atmospheric contribution using this equation is usually not trivial because of the need of, as input, many atmospheric parameters that change in space and time. In particular, one has to know not only the dependence with height of atmospheric temperature, pressure, particle size and type, and humidity, but also how they change spatially and temporally. Some of this information is provided by ground-based measurements and radiosonde ascents, but generally, these measurements are not available in good enough spatial resolution to accurately account for the regional and temporal variability of the atmosphere.

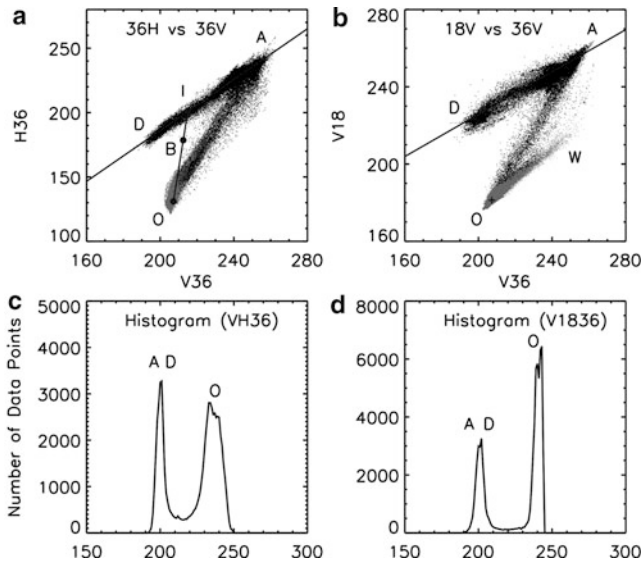
In the polar regions, atmospheric effects are usually not as big a factor as those in the tropics and mid-latitudes because of much lower humidity. The net effect on the retrieval is further minimized through the use of multichannel data and the way the algorithm is structured. Ice concentration, C_I , has been defined as the percentage fraction of sea ice within the field of view of the sensor

and is usually derived using the mixing algorithm $T_{BS} = C_I T_{IS} - (1 - C_I) T_{OS}$, where T_{IS} is the brightness temperature of 100 % ice while T_{OS} is the brightness temperature of 100 % open water. The subscript S is meant to indicate that the parameters represent surface measurements. The sea ice concentration can thus be expressed as the ratio of the difference of brightness temperatures as follows:

$$C_I = \frac{T_{BS} - T_{OS}}{T_{IS} - T_{OS}} \quad (2)$$

According to Equation 1, the brightness temperature of the surface ($T_{BS} = \varepsilon T_S$) is the brightness temperature observed by the satellite (T_B) minus the atmospheric contributions and divided by $e^{-\tau}$. The contributions directly from the atmosphere (second term in Equation 1) within the satellite footprint are approximately the same for the ice-covered and ice-free area, while the downwelling component is significantly smaller and the difference may be negligible. Thus, to a first approximation, the atmospheric effects are either small or may cancel out in Equation 2. Also, the atmospheric opacity, $e^{-\tau}$, which appears in the numerator and denominator in Equation 2, is likely similar over ice-covered, partially ice-covered, and ice-free areas within the ice pack and approximately cancels out as well.

The Bootstrap Algorithm for ice concentration as discussed in Comiso et al. (2003) makes use of an equation similar to that of Equation 2 but with satellite values being substituted for surface values. In particular, a modified Equation 2 is used in which T_{BS} , T_{OS} , and T_{IS} are replaced by T_B , T_O , and T_I , which are variables provided or can be inferred from the satellite multichannel data. In Figure 2, we show two sets of scatter plots using brightness temperature data from AMSR-E: One is the set we call VH36 which makes use of 36 GHz data at both horizontal and vertical polarizations (Figure 2a) and the other is the set V1836 which uses 18 and 37 GHz data at vertical polarization (Figure 2b). The two scatter plots show that the data points from the ice-covered regions are generally confined inside a triangle defined approximately by OAD in the scatter plots. Two main clusters of data points, one along AD and the other along OW, are apparent, and with the aid of high-resolution Landsat and SAR data, we are able to establish that the data points along AD represent consolidated ice while those along OW (gray data points) represent ice-free water surfaces with the relatively calm water near the point O while those in the rough seas near W. The data points between the line AD and O represent ice-covered regions with different concentrations, but data along OA can also represent new ice types of different thicknesses. In Figure 2a, the cluster of points along OW and OA overlaps and is difficult to discriminate, but in Figure 2b, the cluster OW for open water is clearly easier to separate from ice-covered regions within the triangle OAD. Data points in OAD but close to O or the cluster OW are data points near the



Sea Ice Concentration and Extent, Figure 2 Scatter plots of brightness temperatures from AMSR-E of (a) 36 GHz at horizontal polarization versus 36 GHz at vertical polarization; (b) 18 GHz versus 36 GHz both at vertical polarization; (c) frequency histogram of data points perpendicular to the AD line in (a); (d) frequency histogram of data points perpendicular to the AD line in (b).

ice edge, and at some ice concentration values, they are virtually impossible to discriminate from ice-free oceans. The threshold at which such discrimination can be done is usually used to define the ice edge and is about 10 % for AMSR-E and 15 % for SSM/I. We use the 18 and 23 GHz channels as discussed in Comiso et al. (2003) for optimal discrimination and to establish the cutoff threshold for ice-covered regions.

The primary cluster of interest is that along AD, because it is the cluster that is used to determine the tie point (T_I) for consolidated ice. Since the data points in Figure 2a are from the same frequency, the slope of the cluster along AD is approximately equal to 1. This is basically because a unit change in brightness temperature in one channel is reflected as a change of approximately the same magnitude in the other channel. The data points for consolidated ice are confined within a tight linear cluster along AD because incremental changes in temperature or emissivity cause the data points to slide along the same line. The distribution of data points along AD in Figure 2b is not as compact as in Figure 2a because the set of data is from different frequencies and the slope is not equal to 1. The cluster shows a larger range of variability along the 37 GHz data than the 18 GHz data, reflecting the stronger impact of scattering on emissivity for radiation at the shorter wavelength (i.e., 37 GHz). The variability across the AD cluster in Figure 2b is in part associated with ice surface temperature effects since the slope is not unity while changes in temperature cause the data points to shift along a diagonal line.

The scheme behind the algorithms, and in particular the Bootstrap Algorithm, is to determine the tie-points (T_O and T_I) needed to estimate ice concentration from the modified Equation 2 by using the scatter plots in Figure 2a and b. For any given pair of satellite data (T_B) represented by the data point B in the scatter plot in Figure 2a, the corresponding brightness temperature for 100 % ice (T_I) is inferred by extending the line OB such that it intersects the line AD at I. The location of the point of intersection is determined by solving for the common point of the two linear equations representing the lines AD and OB. The brightness temperature, T_I , at the point of intersection corresponds to 100 % ice for the type of sea ice that is represented by the data point at B. We can estimate the ice concentration from the ratio OB/OI which yields the same result as the modified version of Equation 2 (Comiso et al., 2003).

Choosing the right sets of channels to be used in the algorithm is an important consideration and was based on the ability to discriminate ice and liquid water and at the same time meet some resolution requirements. The resolution of a passive microwave sensor is dependent on the size of the antenna and the wavelength of the radiation and since one antenna is used for all the radiometers, the resolution is frequency (or wavelength) dependent. The 6 GHz channels provide the best contrast in the emissivity of ice and water and at the same time the least sensitivity to surface effects, including the snow cover. Unfortunately, this channel has the poorest resolution at about 40 by 60 km for AMSR-E and a lot worse (about 70 by 150 km) for SMMR data. Another consideration was the sensitivity to surface temperature variations which affects the compactness of the cluster of data points along the line AD. After evaluating the different possibilities, it was determined that the VH36 and V1836 sets of channels provide a consistent and reliable ice concentration at a reasonable resolution. It is fortuitous that SMMR, SSM/I, and AMSR-E sensors have these sets of channels, making it possible to have a continuous and consistent data set of sea ice from November 1978 to the present to be generated.

The two sets of channels are utilized as follows. The VH36 set appears to be an important set because of the tight linear clustering of ice data points and the good accuracy provided by this set in estimating sea ice concentration. A histogram of data along an axis perpendicular to the line AD in Figure 2a is shown in Figure 2c, and the peak of the AD line is quite narrow with a standard deviation of about 2.5 °C. The narrow peak means that T_I can be derived at a relatively high accuracy and since T_O is expected to be relatively constant; ice concentration can be estimated at an optimum accuracy of about 3 %. The use solely of the VH36 set, however, is not adequate because some data points that represent 100 % ice, as observed from Landsat or SAR data, are not within the AD cluster in Figure 2a. One reason is that horizontally polarized radiation is much more sensitive to some types of sea ice-covered surfaces (e.g., those with layering) than

the vertically polarized radiation (Matzler et al., 1984). This problem is not a factor with the V1836 set since in this case, a horizontal polarization channel is not utilized and these same problem data points are found within the AD cluster in the V1836 set. The strategy is thus to use the VH36 set for the data points in the AD cluster for this set and the V1836 set for the rest of the data points. The technique thus takes advantage of the strength of the different channels and provides accurate estimates of ice concentration within the ice pack. To generate ice concentration maps, there needs to be a way to detect land areas since sea ice is an ocean parameter and the algorithm is not able to discriminate land from ocean areas. A fixed land mask, based on published land boundaries, has been used for this purpose, but adjustments are made during big calving events. An ocean mask is also used, based on the set of data presented in Figure 2b and as discussed in Comiso et al. (2003) to remove bad ice concentration data in the middle of the open ocean. Erroneous retrievals at the land/ocean boundaries away from the pack are also apparent and are removed by using an enhanced version of the algorithm developed by Cho et al., 1996.

Generally, the cluster along the line AD, especially for the VH36 set, does not change much during the ice season. However, unusual weather and surface conditions could cause the ice cluster to shift in slope and offset. Results from regression fits to the cluster of points along the line AD are used on a daily basis to constantly update the line parameters for AD and account for changes in weather and surface conditions. The impact of spatial changes in surface temperature on the accuracy of the retrieval has also been studied, and in snow-covered regions (which is the dominant type of ice cover), the standard deviation of surface ice temperatures has been observed to be about 2.5 °C. The associated error in the retrieval accuracy when using the V1836 set is significant due to temperature variations but is relatively small (about 3 %). The technique is thus basically robust because much of the temporal and spatial changes in emissivity and temperature in ice-covered regions are taken into account. From comparative studies with high-resolution satellite data, the accuracy has been estimated to be about 5–10 % for most types of thick ice cover during the dry (cold) period. During the melt period, the emissivity becomes more unpredictable especially during onset of melt and when melt ponds start to form. Depending on the fraction of melt pond in each pixel, the error is estimated at 15 % or greater during the period. Also, since new ice data are represented by data points between AD and O, data points with close to 100 % of new ice are retrieved with relatively lower percentage of ice cover, depending on thickness.

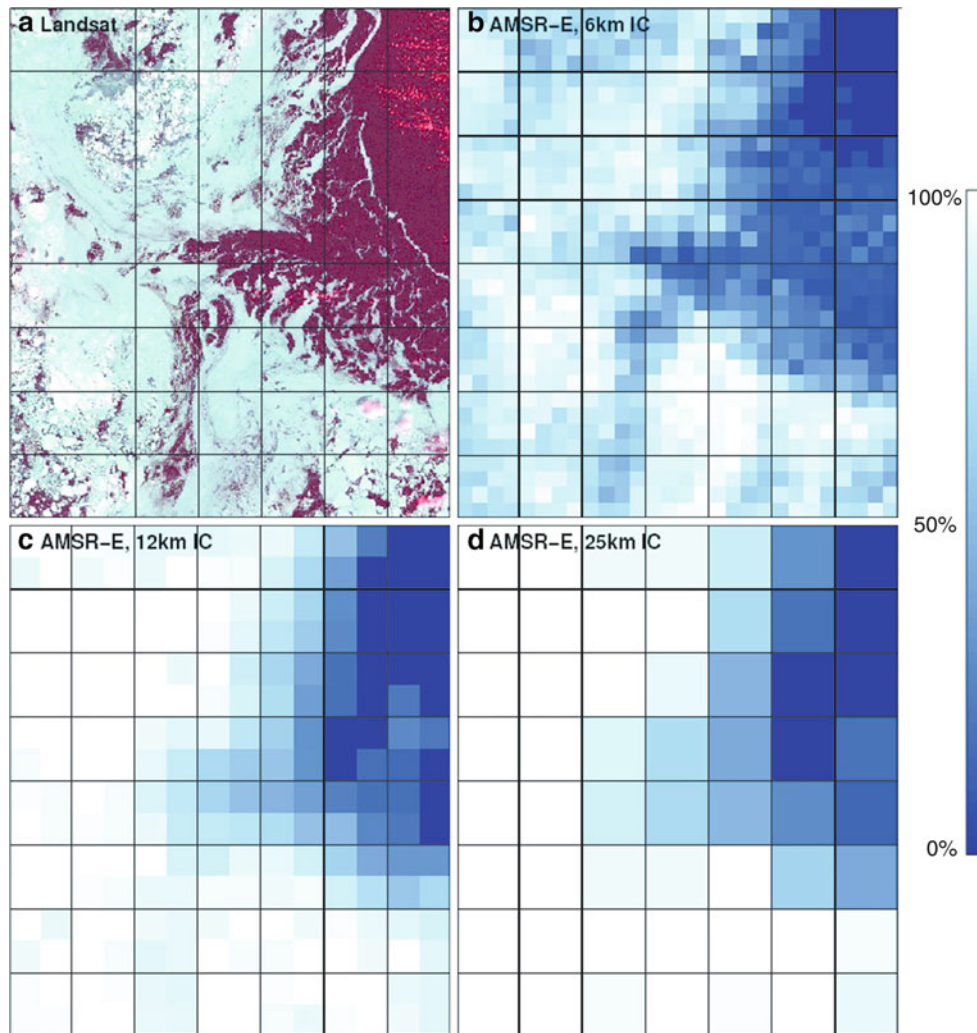
Validation studies have been done using high-resolution data and other techniques (Steffen et al., 1992; Comiso and Steffen, 2001). A sample-comparative analysis, presented in Figure 3, shows a Landsat image during winter (February 11, 2003) in the Sea of Okhotsk and corresponding ice concentration maps at 6.25 km,

12.5 km, and 25 km resolution from AMSR-E data. It is apparent that the 6.25 km data using solely the 89 GHz channels reproduces many of the small-scale features in the Landsat data. The 12.5 km data provide an accurate depiction of the contours along the ice edge, while the 25 km data show a slight smearing of the ice cover and an ice edge that is not as well defined as the 12.5 km data. Quantitative analysis shows very good agreement of Landsat with the passive microwave data, but the 25 km data provided the best correlation with Landsat data, while the 6.25 km data provided the worst correlation. This is in part because of a few hours difference in observation time and a few km ice drift can cause significant mismatch on a pixel-by-pixel basis when the high-resolution AMSR-E data are used.

Sea ice extent and area

The ability to estimate sea ice concentration from satellite data has provided the means to quantify the large-scale characteristics of the global sea ice cover (Gloersen et al., 1992; Bjorgo et al., 1997; Cavalieri et al., 1997; Parkinson et al., 1999; Zwally et al., 2002). To illustrate the seasonality of the sea ice cover, multiyear monthly averages of sea ice concentration from AMSR-E data using data from June 2002 to December 2009 are presented in Figures 4 and 5 for the Northern and Southern Hemispheres, respectively. We use the AMSR-E data because of the relatively high resolution, and they represent the typical ice cover during the contemporary era. The changes in the ice cover with season are shown to be quite different in the two hemispheres with the ice advance and retreat occurring in close proximity to land in the Northern Hemisphere but basically toward the open oceans in the Southern Hemisphere.

In the Northern Hemisphere, the advance of sea ice from September to November occurs at a relatively fast pace in the Arctic basin, Canadian Archipelago, and Baffin Bay. From December to March, the ice advances farther to the south to the Bering Sea and the Okhotsk Sea in the Pacific region and to Hudson Bay, Labrador Sea, and the Gulf of St. Lawrence in the Atlantic region. The sea ice cover reaches its maximum extent in late February or early March and starts retreating after that, with the peripheral seas becoming ice-free in July. The ice cover usually reaches its minimum extent in early to mid-September with the ice that survives the summer being just the relatively thick ice in the Arctic Basin. The Beaufort gyre tends to keep a significant fraction of ice confined in the basin for as long as 7 years, enabling the ice to be relatively thick. The presence of land surrounding the Arctic basin also makes it more likely for ridging and, therefore, thicker ice to occur in the region. Sea ice is also found at relatively low latitudes (about 44 °N) in the Northern Hemisphere in winter mainly because of the influence of extremely cold surface temperatures in the region and adjacent land areas during the period.

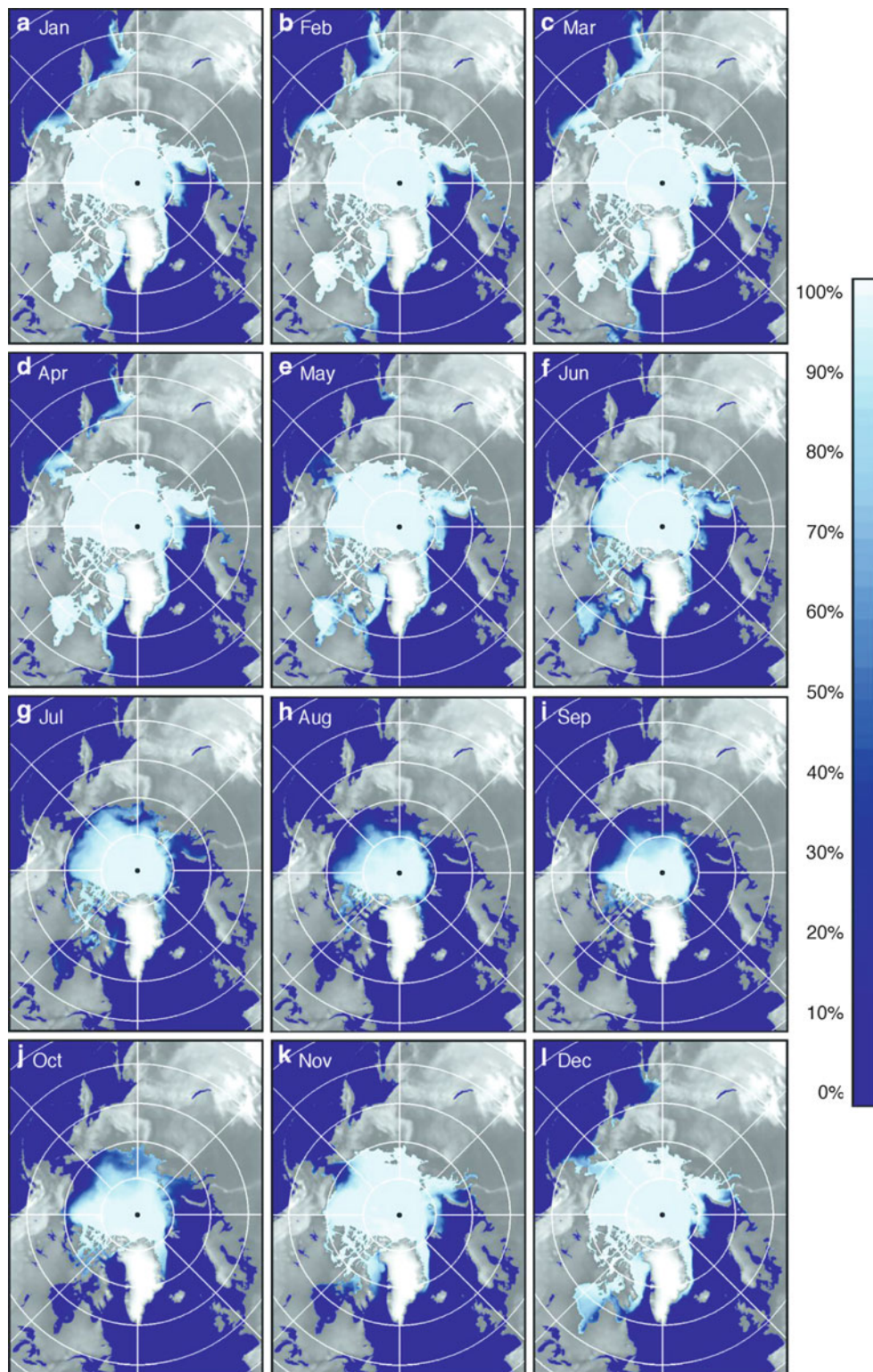


Sea Ice Concentration and Extent, Figure 3 Comparison of (a) Landsat visible channel data with AMSR-E data at (b) 6.25 km, using solely the dual-polarized 89 GHz data; (c) 12.5 km using 18 and 36 GHz data; and SSM/I data at (d) 25 km resolution using 19 and 37 GHz data. The image was taken on February 11, 2003, data at the Sea of Okhotsk.

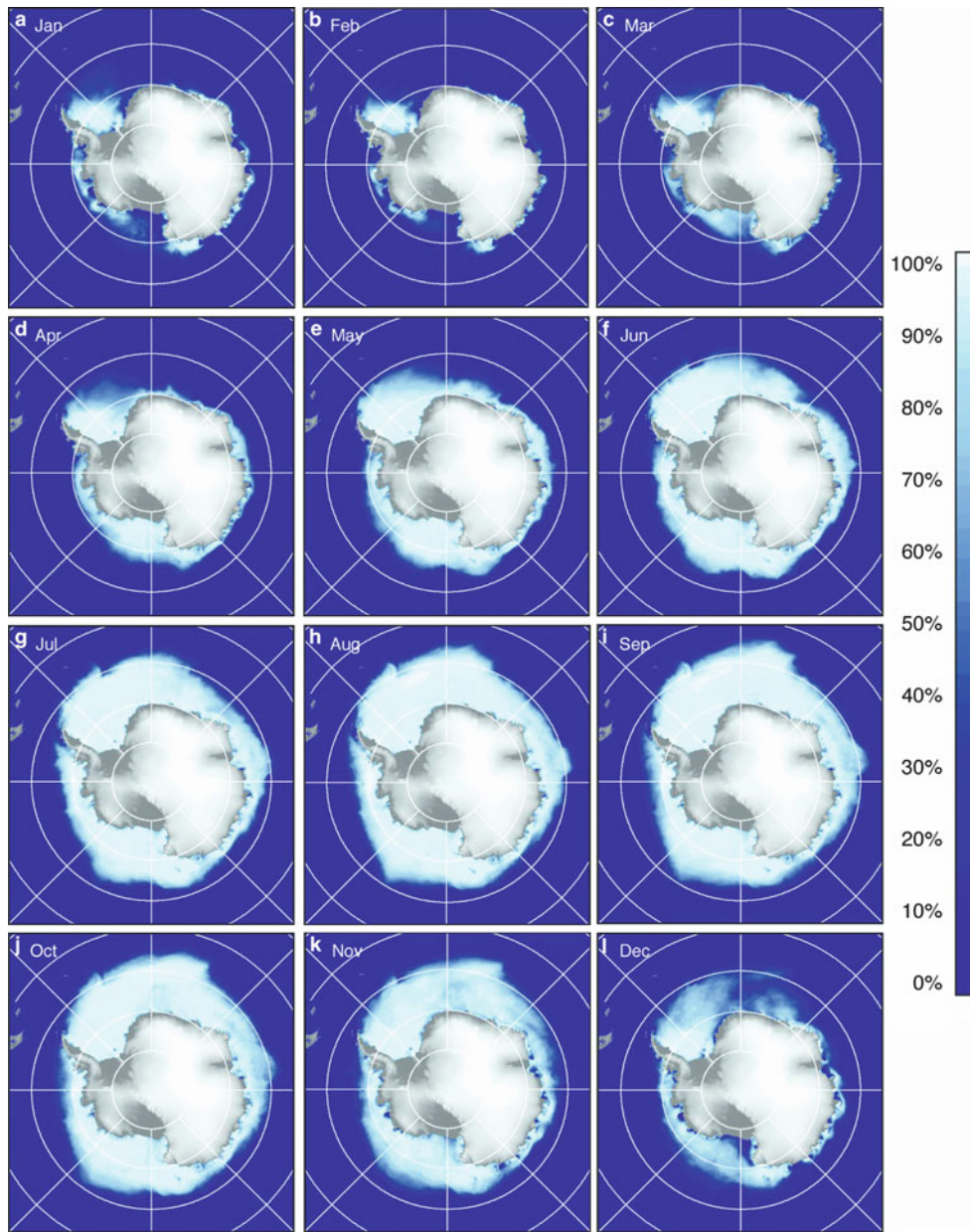
In the Southern Hemisphere, the sea ice cover in the Weddell and Ross Seas advances quickly to the north from March to June (Figure 5), while the entire ice cover becomes almost symmetrical in July, when other regions have caught up advancing to the north, and reaches maximum extent in September or October. The rate of advance is influenced primarily by sea-surface temperature, southerly winds, and ice production in the coastal regions. The multiyear averaging causes some smearing near the ice edges in May and June on account of large interannual changes of ice edge location. In particular, the gradients in ice concentration at the marginal ice zones are less abrupt than those from monthly averages in a single year. The ice cover subsequently declines in November through January reaching minimum extent in late February or March. The abrupt decline in November and December is caused primarily by big waves and intrusion of warm

water from lower latitudes. In the winter, the ice cover is contiguous and appears like a single circumpolar unit that is more extensive than the Antarctic continent. The seas with the most ice cover during the period are the Weddell and Ross Seas which are also sites of large coastal polynyas that are regarded as ice factories. Also, the sea ice cover does not advance much to the north in areas where the location of the Antarctic coastline is at relatively low latitudes (e.g., Western Pacific region between 100°E and 150°E). In the summer, the ice cover is more fragmented and is basically confined to regions from 140°E to 330°E, with the Western Weddell Sea having the most summer ice.

An area of special interest is the marginal ice zone (MIZ) which is usually a region of very intense air-sea-ice interaction and high primary productivity. Sea ice cover images in the Weddell Sea as depicted by MODIS,



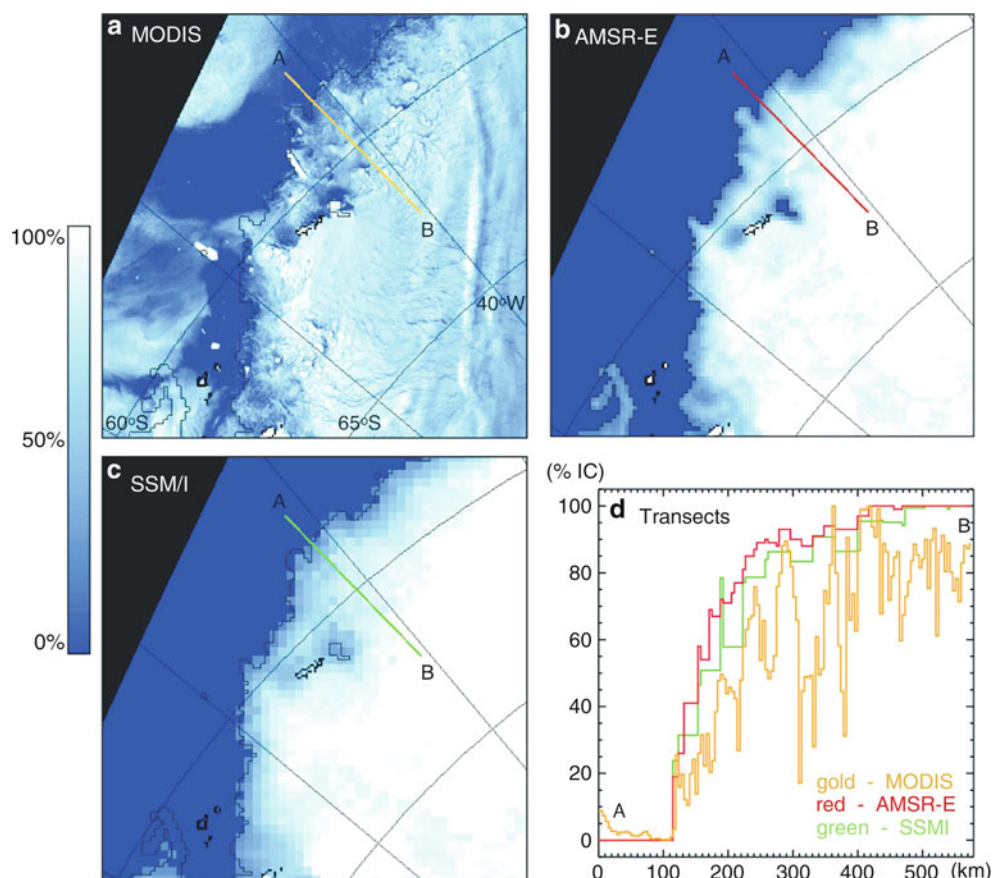
Sea Ice Concentration and Extent, Figure 4 Color-coded multiyear averages (2002–2009) of monthly ice concentration from January to December from AMSR-E data in the Northern Hemisphere.



Sea Ice Concentration and Extent, Figure 5 Color-coded multiyear averages (2002–2009) of monthly ice concentration from January to December from AMSR-E data in the Southern Hemisphere.

AMSR-E, and SSM/I data on September 9, 2002, and mapped to a common projection at a grid size of 250 m, 12.5 km, and 25.0 km, respectively, are presented in [Figure 6a, b, and c](#). The narrow-band albedo data from MODIS have been arbitrarily normalized to obtain a better match with the other images and for convenience in comparative analysis. The MODIS data provide the most detailed characterization of the sea ice cover, while the AMSR-E and SSM/I data show basically similar information but at a lower resolution. It is apparent that resolution can make a significant difference since AMSR-E data

provide a better representation of the mesoscale features of the ice cover than SSM/I data when compared with MODIS data. For example, some of the relatively low concentrations and open water features within the pack that are apparent in the MODIS data are also captured by AMSR-E but not by SSM/I. For a better quantification of the difference in the information provided by the three data sets, plots along a transect defined by a line from A to B, using data from the three data sets, are presented in [Figure 6d](#). The MODIS albedo (in gold) is shown to be relatively low in the open water area (at A) and actually



Sea Ice Concentration and Extent, Figure 6 Sea ice cover in the Weddell Sea on September 9, 2002, as depicted by (a) MODIS albedo data at 250 m resolution; (b) AMSR-E ice concentration data at 12.5 km resolution; and (c) SSM/I ice concentration data at 25 km resolution. (d) Plot of data points across the ice edge and along the line A to B using MODIS (in gold), AMSR-E (in red), and SSM/I (in green) data.

goes down as it approaches the ice edge and then increased to significantly higher values within the pack. Large fluctuations of albedo within the pack are apparent. This is largely due to the spatial variations in ice characteristics and associated changes in the albedo (Allison et al., 1993; Warren, 1984). Plots along the same transect but using ice concentration values from AMSR-E and SSM/I are shown by the red and green lines, respectively. It is apparent that the ice edge locations (i.e., where the ice concentrations increase rapidly), as detected by MODIS, AMSR-E, and the SSM/I data, are consistent. The black line along the ice edge in the AMSR-E image (Figure 6b) is superimposed on the MODIS and SSM/I images for comparative analysis. Generally, the SSM/I image shows an ice edge farther away from the pack than AMSR-E because of coarser resolution. Beyond the ice edge and into the pack, the MODIS data show much more variability in the ice cover than either the AMSR-E or SSM/I data. This is in part associated with the higher resolution that enables the detection of some of the distinct but small ice features within the pack. Some of these features represent different sea ice surfaces which are

represented in the passive microwave data as close to 100 % sea ice. The passive microwave values may indeed be the correct values since many of the features are likely leads that may have intermediate albedo but are completely frozen. The AMSR-E and SSM/I data show a well-defined ice edges and consistent ice cover while MODIS shows large variability that may or may not reflect actual changes in ice concentration.

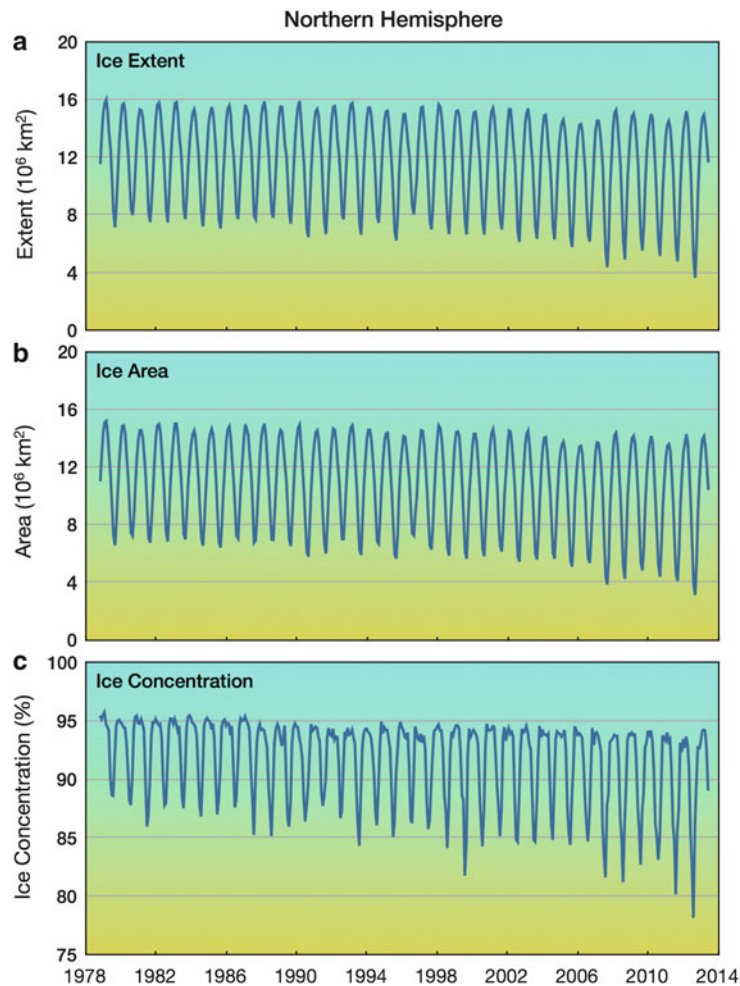
The two parameters that are usually used in the quantitative assessment of the sea ice cover are sea ice extent and ice area. Ice extent refers to the area covered partially or completely by sea ice and is quantified by taking the integrated sum of the areas of data elements (pixels) with at least 15 % ice concentration. The lower limit of 15 % is based on the ability of the sensor to discriminate sea ice from ice-free regions at low concentrations. This can be set lower especially with the more accurate AMSR-E data, but for consistency with historical data and previous studies, we use the same 15 % lower limit. Ice area corresponds to the integrated sum of the products of the area of each pixel and the corresponding ice concentration. In general, ice extent provides information about how far

south (or north) the ice advances during the winter period and how far north (or south) it retreats toward the poles in the summer, while ice area provides the means to estimate the total volume assuming that the average ice thickness is known. Ice area is thus useful in ice variability and mass balance studies.

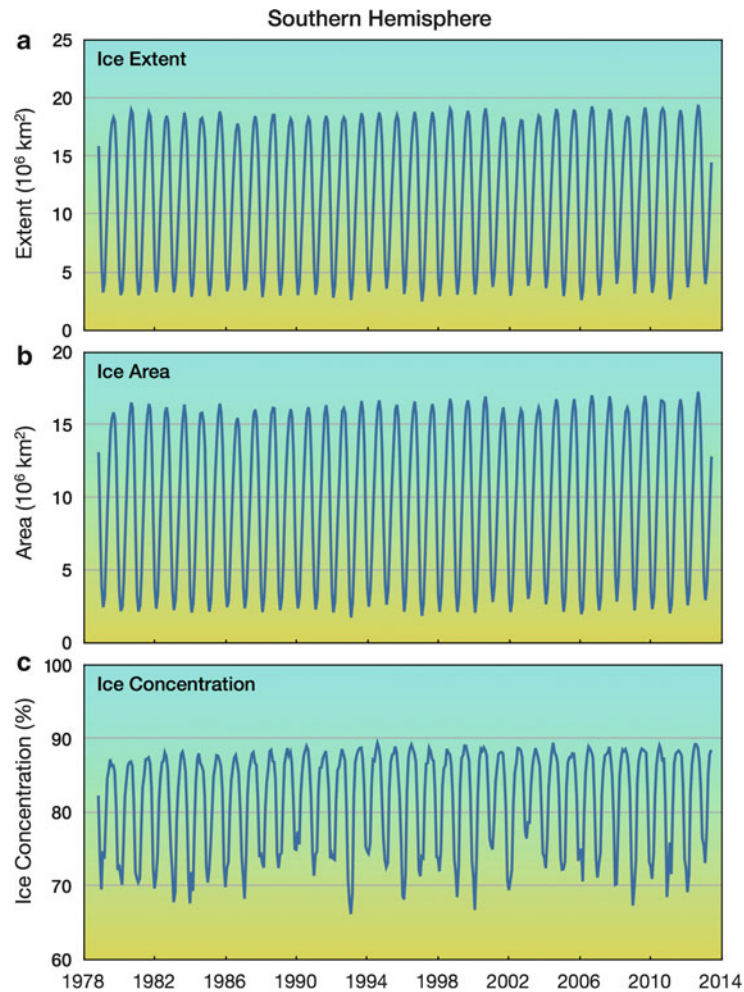
The ability to estimate the ice extent and ice area makes it possible to quantitatively study the seasonal and interannual variability as well as trends in the global sea ice cover. For consistency, the data that have been used for sea ice time series studies are those from SMMR which provided good data from November 1978 to August 1987 and a few SSM/I sensors which have been providing good data from mid-July 1987 to the present. Although the best data set available is the AMSR-E data, the record length is less than 10 years and do not provide a long enough record to document meaningful trends in the ice cover. The AMSR-E data benefits from a higher resolution, wider spectral range, and wider swath than previously available data. The higher resolution enables improved ability to do

comparative studies with other data sets, especially MODIS data that are coincident in time and coverage. The tie-points for AMSR-E were therefore optimized with the help of cloud-free MODIS data. With AMSR-E data as the baseline, SSM/I brightness temperatures were normalized to be consistent with AMSR-E brightness temperatures using normalization parameters derived during the period of overlap. Using basically an identical algorithm and matching tie-points, SSM/I ice concentrations were generated to be consistent with AMSR-E data. A similar process was in turn used to make SMMR data consistent with SSM/I data using normalization parameters derived from analysis of SMMR and SSM/I data during overlap period. Results from detailed analysis of ice concentrations, ice extent, and ice area from this time series have been reported by Comiso and Nishio (2008).

Continuous and consistently derived monthly ice extent, ice area, and ice concentrations from SMMR and SSM/I data for the period from November 1978 to June 2013 are presented in Figures 7 and 8 for the Northern



Sea Ice Concentration and Extent, Figure 7 Monthly ice extents, ice area, and ice concentration in the Northern Hemisphere using SMMR and SSM/I data from November 1978 to June 2013.

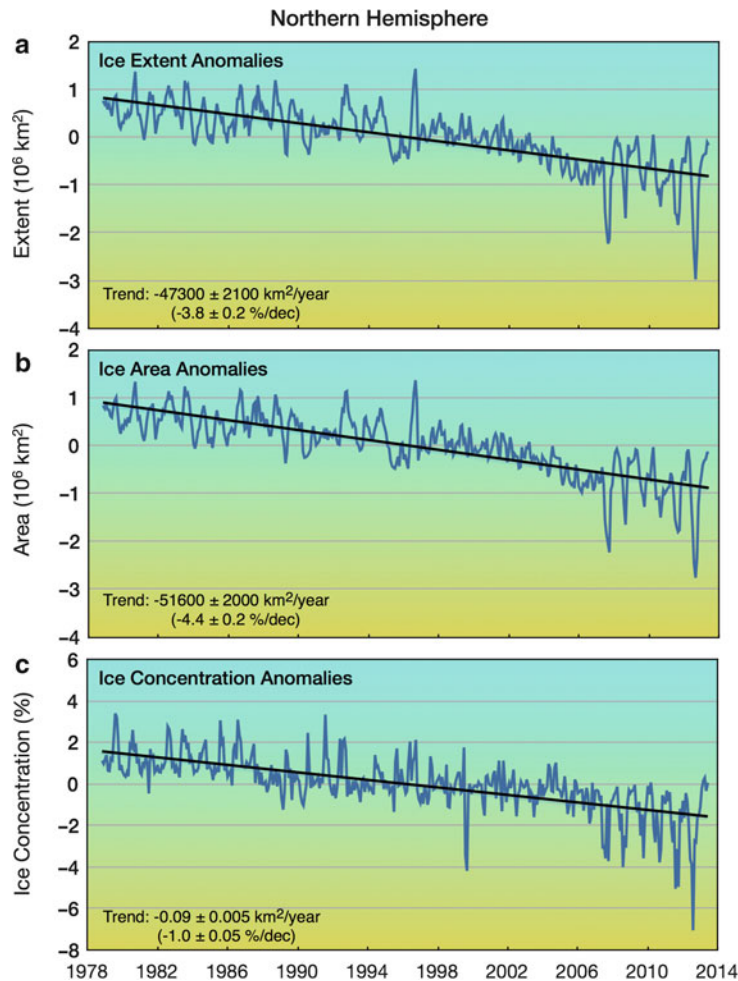


Sea Ice Concentration and Extent, Figure 8 Monthly ice extents, ice area, and ice concentration in the Southern Hemisphere using SMMR and SSM/I data from November 1978 to June 2013.

and Southern Hemispheres. The temporal variations in the monthly averages are dominated by the large seasonality of the ice cover that fluctuates from minimum values in September to maximum values in February in the Northern Hemisphere and minimum values in February to maximum values in September in the Southern Hemisphere. In the Northern Hemisphere, the ice extent fluctuates from about $7.2 \times 10^6 \text{ km}^2$ in the summer to $15.5 \times 10^6 \text{ km}^2$ in the winter in the 1980s, while in the 1990s and 2000s, the range slightly increased as the trend in winter maximum is not able to keep up with the larger downward trend in summer minimum (Figure 7a). Similar changes from the 1980s to 2000s are also shown in Figure 7b for sea ice area. It is intriguing that high values of ice extents in winter are sometimes followed by low values in the summer (and vice versa) as in 1979 and 1990. Such phenomenon suggests that a cold winter that causes the ice extent to reach abnormally high values does not necessarily lead to more ice surviving the summer, although the ice

cover would be expected to be relatively thick. It also means that other parameters (e.g., wind circulation and percentage of ice exported to the Atlantic Ocean through Fram Strait) have to be considered when making projections about the ice cover for a subsequent season. It is also interesting that while the minimum ice extent and area remain reasonably constant from 1979 to 1990, larger variability occurred thereafter and appears to be significantly lower than previous values starting with 1997. Some recoveries are apparent in the summers of 1992, 1994, and 1996, but since 2002, such recovery has been relatively minor and the five lowest extents during the satellite era occurred in 2007 to 2011. Not shown is the new record minimum ice extent in September 2012.

Monthly extents and ice areas in the Southern Hemisphere, as derived from SMMR and SSM/I data (Figure 8), show an even more seasonal ice cover than that of the Northern Hemisphere. Minimum ice extents and ice areas usually occur in February, while maximum ice extents



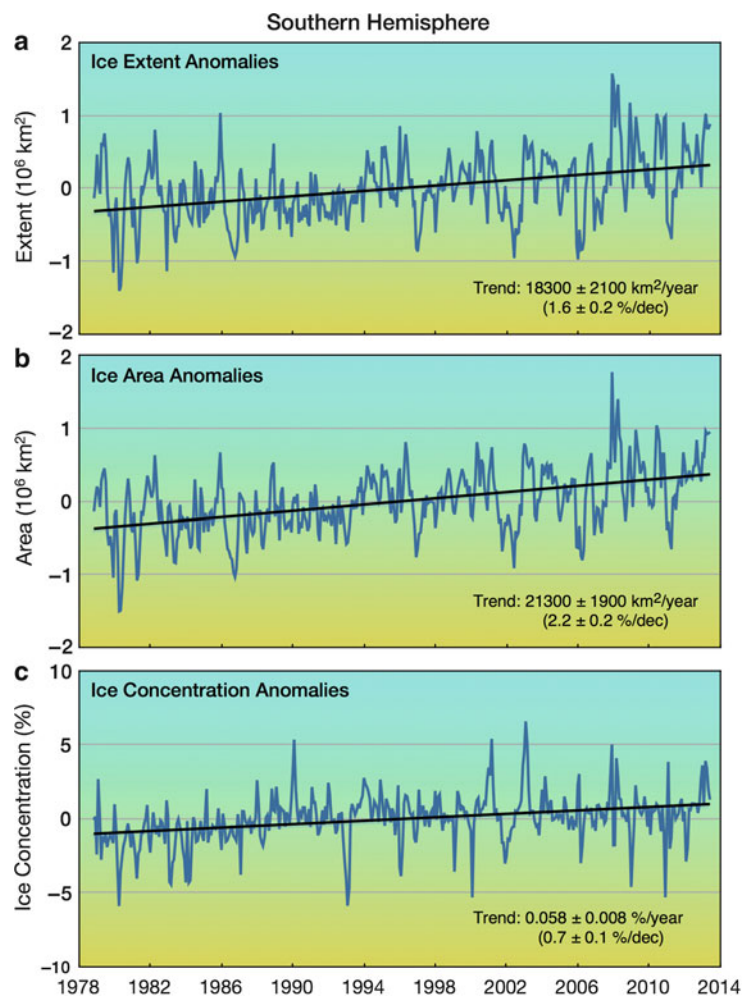
Sea Ice Concentration and Extent, Figure 9 Monthly anomalies of (a) ice extents; (b) ice area; and (c) ice concentration in Northern Hemisphere from November 1978 to June 2013. The bold black lines through the data points are the results of linear regression.

and ice areas occur in September. This means that the growth period takes a longer time than the melt period in the Southern Hemisphere. The maximum and minimum extents and areas go through interannual fluctuations, but they appear relatively stable. It is interesting to note that the maximum values increased from 2003 to 2006 while the minimum values decreased during the same period. In the following 2 years, the reverse happened, but the pattern started to go in the opposite direction in 2009. Such phenomenon suggests that the influence of conditioning from the ocean or previous ice patterns is not strong and that the influence of atmospheric circulation on the interannual variability of the ice cover is likely more important.

Monthly anomalies and trends in ice extent, area, and ice concentration

To assess interannual trends in the ice cover, we use monthly anomalies as has been done previously (Parkinson et al., 1999; Zwally et al., 2002) in order

to minimize the uncertainties associated with the large seasonal variations. The monthly anomalies were estimated by subtracting from each monthly average the corresponding monthly climatological average which in our case is the average of all data for the months from November 1978 to June 2013. The monthly anomalies for the ice extent, ice area, and ice concentration in the Northern Hemisphere are presented in Figure 9 using the enhanced data from SMMR and SSM/I. Interannual fluctuations in extent and area are apparent but are usually less than one million square kilometers except for a few occasions as in 1996, 2007, 2008, 2011 and 2012. The latter correspond to a significant increase in the September ice cover in 1996 and a considerable drop in the September ice cover in 2007, 2008, 2011 and 2012. There was also a significant drop in the ice cover in September of 2009 and 2010 but not as much as the other two. The time series shows that the trends in ice extent and ice area in the Northern Hemisphere are -3.8 ± 0.2 and $-4.4 \pm 0.2 \text{ \%/dec}$ per decade, respectively, the difference associated with



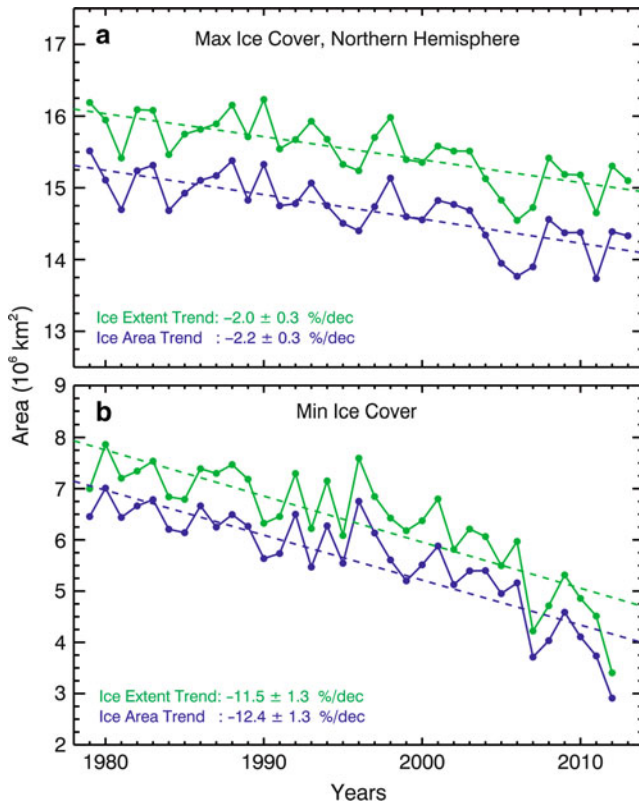
Sea Ice Concentration and Extent, Figure 10 Monthly anomalies of (a) ice extents; (b) ice area; and (c) ice concentration in Southern Hemisphere from November 1978 to June 2013. The **bold black lines** through the data points are the results of linear regression.

the negative trend in ice concentration of $-1.0 \pm 0.1 \text{ \%/decade}$. The decline in ice concentration is relatively minor and can be caused by changes in wind patterns that could cause more divergence in the region. The ice concentration anomalies also show significant drops every year from 1999 to 2012 that may be associated with more storms in the summer during these years.

Similar plots for the anomalies in ice extent, ice area, and ice concentrations but for the Southern Hemisphere are presented in Figure 10. The data show much larger interannual fluctuations in ice extent and ice area in the 1980s than those for the corresponding period at the Northern Hemisphere. Significant variability is also apparent from the year 2000. It is intriguing that anomalously high values occurred in 2008 which also coincide with a rapid recovery in the ice cover in the winter of early 2008 in the Northern Hemisphere from its record of low values in September 2007. This phenomenon is likely

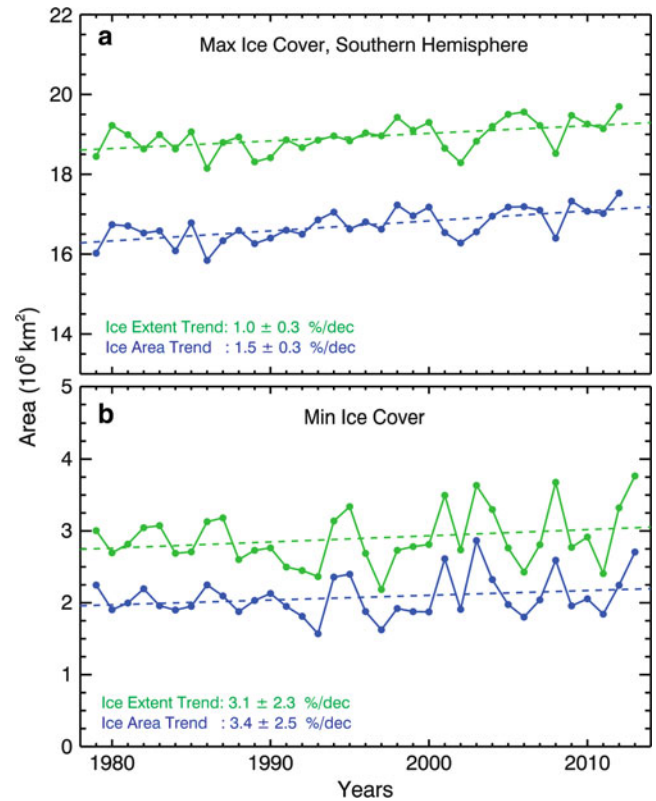
due to an apparent global cooling associated with the La Niña in 2007. The trends for extent and area are shown to be positive at 1.6 ± 0.2 and $2.2 \pm 0.2 \text{ \%/decade}$, respectively, which are significantly higher than previous estimates (e.g., Zwally et al., 2002). Again, the difference in the trends for ice extent and ice area is in part associated with the slight positive trend in ice concentration which is estimated to be about $0.7 \pm 0.1 \text{ \%/decade}$.

The relatively good time resolution of the data also enables the study of the variability of the maximum and minimum extents and areas of the ice cover for each year. The identification of the actual day when the maximum and minimum extents occurred for each year is inferred from the daily data using 5 day running averages, and the plots of maximum and minimum ice extents and areas in the Northern and Southern Hemispheres for the period 1979 to 2013 are presented in Figures 11 and 12. The data provide different values than those from monthly averages



Sea Ice Concentration and Extent, Figure 11 Yearly (a) maximum and (b) minimum extents and ice area of the ice cover in the Northern Hemisphere. Dashed lines correspond to linear regression fit to the data, and trends with statistical errors are indicated.

because of daily changes that are not reflected in the latter. Large interannual variability is apparent, especially in the minimum ice extent and ice area, but the years of high variability do not coincide for the two hemispheres. In the Northern Hemisphere, the trends in maximum extent and area are -2.2 ± 0.3 and $-2.2 \pm 0.4 \text{ \%/dec}$, respectively, while the corresponding values in the Southern Hemisphere are 1.0 ± 0.3 and $1.5 \pm 0.3 \text{ \%/dec}$, respectively. The signs of the trends in the two hemispheres are different, as expected from previous analysis, but the trends are relatively low. The trends in minimum extent and area in the Northern Hemisphere are more dramatic at -11.5 ± 1.6 and $-12.4 \pm 1.6 \text{ \%/dec}$, respectively. Similar results have been reported in several publications (e.g., Comiso, 2002; Comiso et al., 2008), suggesting a rapid decline and a possible disappearance of the Arctic perennial ice cover in a foreseeable future. On the other hand, the trends in minimum extent and area in the Southern Hemisphere are positive at 3.1 ± 2.3 and $3.4 \pm 2.5 \text{ \%/dec}$, respectively, but the relatively large statistical errors suggest that these trends may not be significant.



Sea Ice Concentration and Extent, Figure 12 Yearly (a) maximum and (b) minimum ice extent and ice area of the ice cover in the Southern Hemisphere. Dashed lines correspond to linear regression fit to the data, and trends with statistical errors are indicated.

Discussion and conclusions

Passive microwave satellite data currently provide the only means to obtain a long-term and consistent record of the global sea ice cover. Through the use of an algorithm that utilizes two sets of two-channel data, ice concentration maps in both polar regions have been derived with accuracies that range from about 5 % to 10 % in highly packed ice during dry and cold conditions, around 10–15 % near the marginal ice zones and polynya areas, and 15 % or greater during the peak of the summer when the ice surface is wet and significant meltponding occurs. Consistently derived ice concentration maps at a resolution of 25 km, using SMMR and SSM/I data, have been generated and provide more than three decades of historical data that are suitable for the study of the interannual variability and trends in the ice cover. Meanwhile, newer and more accurate data sets from AMSR-E and AMSR2 with a resolution of 12 km are being generated. This offers the potential of improved characterization of the ice cover and more accurate trend analysis. However, combining AMSR-E with the historical data set for long-term analysis is not trivial because of a potential bias

associated with the different resolutions of the different sensors that causes the retrieval of slightly different extents from AMSR-E and SSM/I data during periods of overlap.

The current time series of ice extent and area data reveal significant trends in the ice cover in the Arctic but counter-intuitive trends in the Antarctic. The results of our regression analysis on monthly ice extent and area anomalies yielded trends of -3.8 ± 0.2 and -4.4 ± 0.2 % per decade, respectively, for the Northern Hemisphere. The corresponding values for the Southern Hemisphere are 1.6 ± 0.2 and 2.2 ± 0.2 % per decade. The trends are 2–3 times higher in the Arctic than the Antarctic, but the signs are opposite. The trends in the yearly maximum extent and area are -2.0 ± 0.3 and -2.2 ± 0.4 % per decade, respectively, in the Northern Hemisphere and 1.0 ± 0.3 and 1.5 ± 0.3 % per decade in the Southern Hemisphere. These trends are significant but relatively small. The trends in the yearly minimum ice extent and area in the Northern Hemisphere are more dramatic at -11.5 ± 1.6 and -12.4 ± 1.6 % per decade, respectively. These results are very important since they suggest that the perennial ice, which is the mainstay of the Arctic sea ice cover, could disappear in the foreseeable future. Trends for the minimum extents and area in the Southern Hemisphere are 3.1 ± 2.3 and 3.4 ± 2.5 which are relatively small and statistically insignificant. Sea ice thus provides a strong climate signal in the Arctic but a weak and poorly understood signal in the Antarctic. Although new insights into the slight positive trend in the Antarctic have been reported, including the possible impact of the ozone hole, further studies are needed to better interpret these results and understand their significance.

Bibliography

- Allison, I., Brandt, R. E., and Warren, S. G., 1993. East Antarctic sea ice: Albedo, thickness distribution, and snow cover. *Journal of Geophysical Research*, **98**(C7), 12417–12429.
- Bjorgo, E., Johannessen, O. M., and Miles, M. W., 1997. Analysis of merged SSMR-SSM/I time series of Arctic and Antarctic sea ice parameters 1978–1995. *Geophysical Research Letters*, **24**, 413–416.
- Cavalieri, D. J., Gloersen, P., and Campbell, W. J., 1984. Determination of sea ice parameters with the Nimbus7 SMMR. *Journal of Geophysical Research*, **89**, 5355–5369.
- Cavalieri, D. J., Gloersen, P., Parkinson, C., Comiso, J., and Zwally, H. J., 1997. Observed hemispheric asymmetry in global sea ice changes. *Science*, **278**(7), 1104–11067.
- Cho, K., Sasaki, N., Shimoda, H., Sakata, T., and Nishio, F., 1996. Evaluation and improvement of SSM/I sea ice concentration algorithms for the Sea of Okhotsk. *Journal Remote Sensing of Japan*, **16**(2), 47–58.
- Comiso, J. C., 1986. Characteristics of winter sea ice from satellite multispectral microwave observations. *Journal of Geophysical Research*, **91**(C1), 975–994.
- Comiso, J. C., 2002. A rapidly declining Arctic perennial ice cover. *Geophysical Research Letters*, **29**(20), 15656, doi:10.1029/2002GL015650.
- Comiso, J. C., 2010. Variability and trends of the global sea ice cover, Chapter 6. In Thomas, D., and Dieckmann, G. (eds.), *Sea Ice – An Introduction to Its Physics, Biology, Chemistry, and Geology*. Oxford: Wiley/Blackwell, pp. 205–246.
- Comiso, J. C., and Nishio, F., 2008. Trends in the sea ice cover using enhanced and compatible AMSR-E, SSM/I, and SMMR data. *Journal of Geophysical Research*, **113**, C02S07, doi:10.1029/2007JC004257.
- Comiso, J. C., and Parkinson, C. L., 2008. Arctic sea ice parameters from AMSR-E using two techniques, and comparisons with sea ice from SSM/I. *Journal of Geophysical Research*, **113**, C02S05, doi:10.1029/2007JC004255.
- Comiso, J. C., and Steffen, K., 2001. Studies of Antarctic sea ice concentrations from satellite data and their applications. *Journal of Geophysical Research*, **106**(C12), 31361–31385.
- Comiso, J. C., Cavalieri, D. J., and Markus, T., 2003. Sea ice concentration, ice temperature, and snow depth, using AMSR-E data. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(2), 243–252.
- Comiso, J. C., Parkinson, C. L., Gersten, R., and Stock, L., 2008. Accelerated decline in the Arctic sea ice cover. *Geophysical Research Letters*, **35**, L01703, doi:10.1029/2007GL031972.
- Gloersen, P., Campbell, W., Cavalieri, D., Comiso, J., Parkinson, C., and Zwally, H. J., 1992. Arctic and Antarctic sea ice, 1978–1987: satellite passive microwave observations and analysis. *NASA Special Publication*, **511**, 290.
- Gordon, A. L., and Comiso, J. C., 1988. Polynyas in the Southern Ocean. *Scientific American*, **256**, 90–97.
- Grenfell, T. C., 1992. Surface-based passive microwave studies of multiyear ice. *Journal of Geophysical Research*, **97**(C3), 3485–3501.
- Holland, M. M., and Bitz, C. M., 2003. Polar amplification of climate change in coupled models. *Climate Dynamics*, **21**, 221–232.
- Kumerow, C., 1993. On the accuracy of the Eddington approximation for radiative transfer in the microwave frequencies. *Journal of Geophysical Research*, **98**, 2757–2765.
- Lubin, D., and Massom, R., 2006. *Polar Remote Sensing*. Berlin: Springer. Atmosphere and Oceans, Vol. 1.
- Martin, S., 2004. *An Introduction to Ocean Remote Sensing*. Cambridge: Cambridge University.
- Martin, S., Drucker, R. S., and Kwok, R., 2007. The areas and ice production of western and central Ross Sea polynyas, 1992–2002, and their relation to the B-15 and C-19 iceberg events of 2000 and 2002. *Journal of Marine Systems*, **68**, 201–214.
- Massom, R., 1991. *Satellite Remote Sensing of Polar Regions*. London: Belhaven Press.
- Matzler, C., Ramseier, R. O., and Svendsen, E., 1984. Polarization effects in sea ice signatures. *IEEE Journal of Oceanic Engineering*, **OE-9**, 333–338.
- Parkinson, C. L., and Comiso, J. C., 2008. Antarctic sea ice from AMSR-E from two algorithms and comparisons with sea ice from SSM/I. *Journal of Geophysical Research*, **113**, C02S06, doi:10.1029/2007JC004253.
- Parkinson, C. L., Cavalieri, D. J., Gloersen, P., Zwally, H. Z., and Comiso, J. C., 1999. Arctic sea ice extents, areas, and trends, 1978–1996. *Journal of Geophysical Research*, **104**(C9), 20837–20856.
- Smith, Jr. W., and Comiso, J. C., 2008. The influence of sea ice primary production in the Southern Ocean: a satellite perspective. *Journal of Geophysical Research*, **113**, C05S93, doi:10.1029/2007JC004251.
- Steffen, K., Cavalieri, D. J., Comiso, J. C., St. Germain, K., Gloersen, P., Key, J., and Rubinstein, I., 1992. The estimation of geophysical parameters using Passive Microwave Algorithms, Chapter 10. In Carsey, F. (ed.), *Microwave Remote Sensing of Sea Ice*. Washington, DC: American Geophysical Union, pp. 201–231.
- Stroeve, J. C., Serreze, M. C., Fetterer, F., Arbetter, T., Meier, M., Maslanik, J., and Knowles, K., 2004. Tracking the Arctic's

- shrinking ice cover: another extreme September minimum in 2004. *Geophysical Research Letters*, 32, doi:10.1029/2004GL021810.
- Svendsen, E., Kloster, K., Farelly, B., Johannessen, O. M., Johannessen, J. A., Campbell, W. J., Gloersen, P., Cavalieri, D., and Matzler, C., 1983. Norwegian remote sensing experiment: evaluation of the Nimbus 7 Scanning Microwave Radiometer for sea ice research. *Journal of Geophysical Research*, 88(C5), 2755–2769.
- Swift, C. T., Fedor, L. S., and Ramseier, R. O., 1985. An algorithm to measure sea ice concentration with microwave radiometers. *Journal of Geophysical Research*, 90(C1), 1087–1099.
- Tucker, W. B., Perovich, D. K., and Gow, A. J., 1992. Physical properties of sea ice relevant to remote sensing, 2. In Carsey, F. (ed.), *Microwave Remote Sensing of Sea Ice*. Washington, DC: American Geophysical Union, pp. 9–28.
- Turner, J., Comiso, J. C., Marshall, G. J., Connolley, W. M., Lachlan-Cope, T. A., Bracegirdle, T., Wang, Z., Meredith, M., and Maksym, T., 2009. Antarctic sea ice extent increases as a result of anthropogenic activity. *Geophysical Research Letters*, 36, L08502, doi:10.1029/2009GL037524.
- Vant, M. R., Gray, R. B., Ramseier, R. O., and Makios, V., 1974. Dielectric properties of fresh and sea ice at 10 and 35 GHz. *Journal of Applied Physics*, 45(11), 4712–4717.
- Wadhams, P., 2002. *Ice in the Ocean*. London: Gordon and Breach Science Publishers.
- Warren, S., 1984. Optical constants of ice from the ultraviolet to the microwave. *Applied Optics*, 23, 1206–1225.
- Weeks, W. F., and Ackley, S. F., 1986. The growth, structure and properties of sea ice. In Unterstiener, N. (ed.), *The Geophysics of Sea Ice*. New York: Plenum. NATO ASI Ser. B, Vol. 146, pp. 9–164.
- Worby, A. P., and Comiso, J. C., 2004. Studies of Antarctic sea ice edge and ice extent from satellite and ship observations. *Remote Sensing of Environment*, 92(1), 98–111.
- Zhang, J., 2007. Increasing Antarctic sea ice under warming atmospheric and oceanic conditions. *Journal of Climate*, 20, 2515–2529.
- Zwally, H. J., Comiso, J. C., Parkinson, C. L., Campbell, W. J., Carsey, F. D., and Gloersen, P., 1983. Antarctic sea ice 1973–1976 from satellite passive microwave observations. *NASA Special Publication*, 459, 1983.
- Zwally, H. J., Comiso, J. C., Parkinson, C. L., Cavalieri, D. J., and Gloersen, P., 2002. Variability of the Antarctic sea ice cover. *Journal of Geophysical Research*, 107(C5), 1029–1047.
- Mean sea level.* The sea level at a single location averaged over short-term variations such as tides and storms.
- Relative sea level rise.* Change in mean sea level at a location along a coast, as would be measured by a tide gauge. Relative sea level change could be caused by changes in the ocean or by a local uplift or subsidence of the land.
- Steric sea level.* A change in sea level caused by a change in the density of the seawater.
- Thermosteric sea level.* The component of steric sea level caused by changes in ocean temperature.
- Halosteric sea level.* The component of steric sea level caused by changes in the salinity of seawater, almost always due to the addition or removal of freshwater.
- Eustatic sea level.* Although the term “eustatic” is antiquated, it is still occasionally used in the literature. Originally, the term was applied by geologists to distinguish globally averaged sea level from local or relative sea level. On geologic timescales, the primary cause of sea level change is changes in the amount of freshwater in the ocean, and steric changes were often ignored. More recently, the term has been used in some oceanographic literature to refer explicitly to the freshwater, or mass, component of globally averaged sea level rise.

Introduction

Over the past decade, popular as well as scientific interest in sea level rise has been fueled by the gradual realization that much of the modern-day rise is caused by human-induced warming of Earth's climate. Over the next century, sea level rise is expected to incur substantial socioeconomic costs due to inundation, increased likelihood of flooding of low-lying areas, loss of coastal wetlands, and many other impacts (IPCC, 2007a). Although the probability is high that sea level will continue rise as the Earth warms, projections of the amount of rise over the next century remain very uncertain (IPCC, 2007b). Nevertheless, with recent observations from satellites, a new in situ observing network for temperature and salinity, tide gauges, and ongoing efforts to reconstruct the sea level record of the past 20,000 years, great strides have been made in understanding the causes and context of modern-day sea level rise (Church et al., 2008).

Holocene

At the peak of the last glacial period about 20,000 years ago, the oceans stood around 120 m below their modern-day levels, and vast ice sheets covered much of North America and Europe. Over the next 10,000 years, the Earth's climate warmed by 4–7 °C and the massive glaciers began to melt, causing rapid sea level rise. Between 8,000 and 6,000 years ago, the rate of rise slowed, reaching modern-day levels about 3,000 years ago (Jansen et al., 2007). Since that time the level of the oceans is thought to have been fairly stable, up until the last century or two. The exact timing of the onset of modern-day sea

Cross-references

[Cryosphere, Measurements and Applications](#)
[Ice Sheets and Ice Volume](#)
[Microwave Dielectric Properties of Materials](#)
[Microwave Radiometers](#)
[Microwave Surface Scattering and Emission](#)
[Sea Ice Albedo](#)

SEA LEVEL RISE

Josh Willis
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Definition

Sea level rise. A general term referring to an increase in the height of the ocean over time.

level rise is not known, but the rate of rise over the past 100 years is large compared to average rates of the past few 1,000 years.

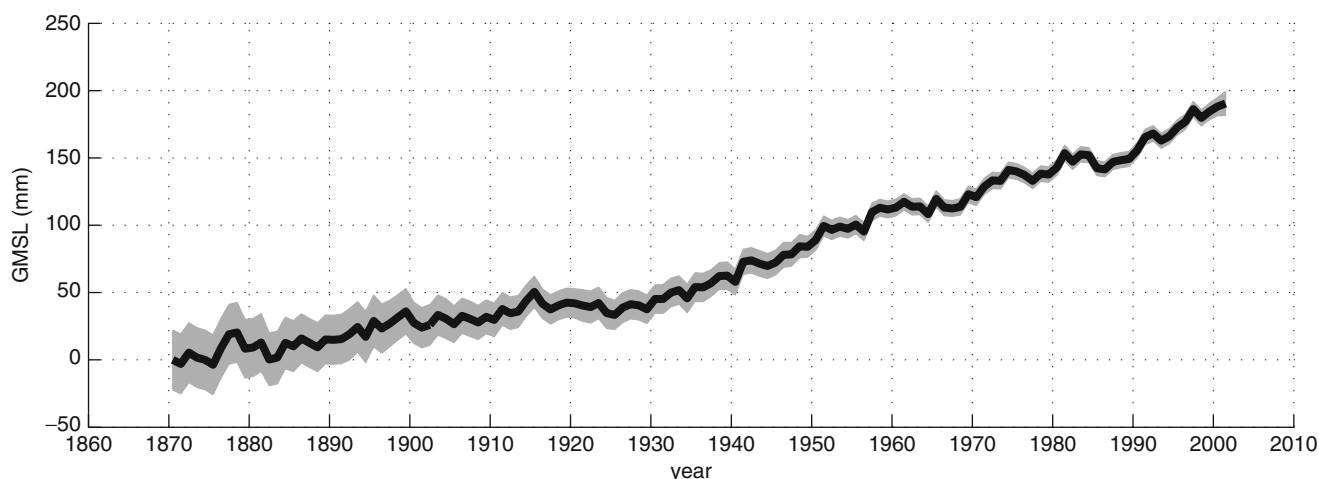
Tide gauge record

Ultimately, what matters most to societies and ecosystems that occupy coastal zones is the local variation in relative sea level or a change in the height of the ocean level relative to that of the land. Traditionally, such changes are measured by tide gauges. Despite the impressive advances of in situ and satellite-based ocean measurements of the past two decades, tide gauge measurements form the cornerstone of the instrumental record of [sea level rise](#) and remain an extremely important source of data. Tide gauge records stretching back to the 1870s ([Figure 1](#)) show that sea level rose by about 20 cm between then and the year 2000, giving an average rate of about 1.8 mm per year over that period (Church and White, 2006). However, tide gauges have several limitations, and chief among these is the fact that sea level rise does not occur uniformly. Changes in ocean circulation may cause local fluctuations that last years or even decades. On longer timescales, changes in the Earth's gravity field also reshape the ocean's surface and prevent spatially uniform sea level rise. Such changes can be driven by changes in the Earth's crust due to processes such as postglacial rebound or even melting of the large ice sheets in Greenland and Antarctica (Mitrovica et al., 2001). Tide gauge measurements also contain signals caused by vertical motion of the land. Although gravitation effects and vertical land motion are often modeled and removed, their presence in the tide gauge data complicates the problem of estimating globally averaged sea level rise (Cazenave and Nerem, 2005).

In situ measurements

On shorter than geologic timescales, the total salt content of the oceans is constant. Increases in ocean volume can therefore be attributed to a net input of either heat or freshwater. Heat-induced volume changes are quantified by globally averaged thermosteric sea level, or the volume-averaged seawater density change caused by thermal expansion. Because of their enormous heat capacity, the oceans are by far the largest reservoirs for the storage of excess heat in the climate system. Over the past 50 years, over 90 % of the excess heat retained by the Earth due to anthropogenic forcing was stored in the world's oceans (Bindoff et al., 2007). For this reason, observations of thermosteric sea level are of particular importance for understanding global warming and the impact of human activities on the climate system.

Comprehensive, high-quality measurements of upper oceans temperature and salinity did not exist prior to the widespread deployment of autonomous floats beginning in 2000 as part of the Argo Project (<http://www.argo.ucsd.edu>). Argo floats now provide global measurements of temperature and salinity in the upper half (about 2 km) of the world oceans. Prior to this, broadscale measurements of temperature were made over much of the oceans, but many were not of sufficient accuracy for climate research. Recently these observations have been recalibrated in order to make estimates of ocean warming thermosteric sea level rise over the past half century (Domingues et al., 2008). These estimates suggest that about one-quarter of the total [sea level rise](#) signal over the past 40 years was caused by thermal expansion in the upper ocean. It is likely that warming in the deep oceans also contributes to sea level rise, but observations of the deep ocean are so sparse that deep warming is difficult to assess in a globally averaged sense.



Sea Level Rise, Figure 1 Globally averaged sea level rise shows a small but significant acceleration between the 1870s and 2000 (Plot is adapted from Church and White (2006) and is based on tide gauge data and spatial patterns from satellite observations of sea level variability).

Satellite era

Beginning with the launch of TOPEX/Poseidon in 1992, satellite altimeters have revolutionized the study of sea level rise by providing nearly global observations of sea level change with local accuracies of a few centimeters. Because ocean currents cause a tilt in the sea surface, changes in ocean circulation can also cause local or regional changes in sea level. The altimeter data showed that such changes often cause local sea level to rise or fall by tens of centimeters over the course of a single year (Figure 2). Furthermore, in situ observations like those of the Argo Project have shown that most of these sea level signals can be attributed to steric, or density-related, changes in the upper oceans (Figure 2). These data have helped to quantify the magnitude and frequency of variability present in all sea level records and to determine the amount of spatial and temporal averaging needed to estimate changes in globally averaged sea level.

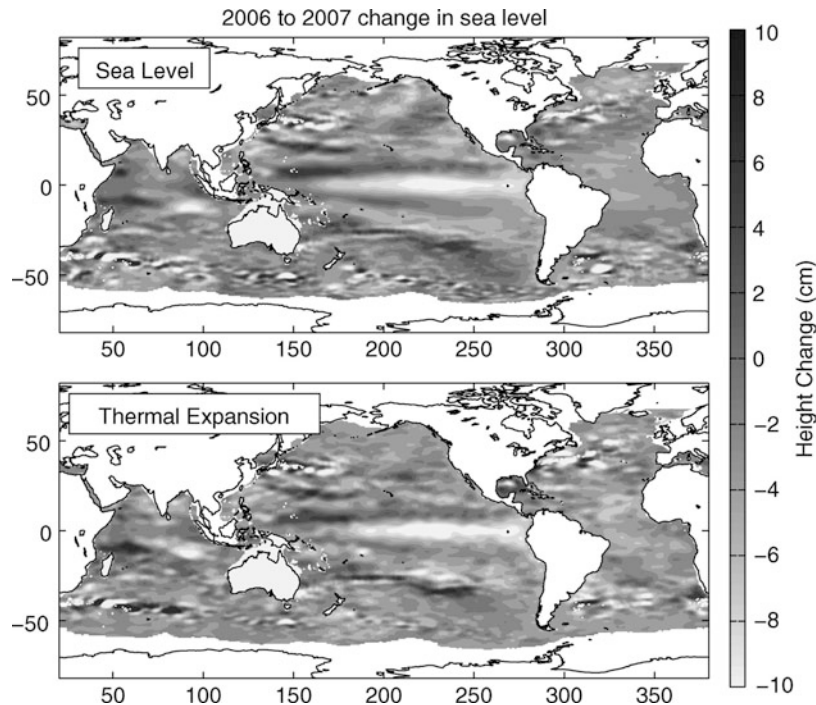
Although meticulous calibration and comparisons with tide gauges are still necessary, altimeters now provide the most accurate ongoing record of global sea level rise. Between 1993 and 2008, sea level rose fairly steadily at a rate of about 3 mm per year. This is larger than the 1.8 mm per year average over the last 100 years, and recent studies have attributed this to an acceleration of sea level rise over the course of the 130 year tide gauge record (Church et al., 2006).

Estimates of the contribution to sea level rise from glacier and ice sheet mass loss as well as groundwater use and

sequestration remain highly uncertain over the 130 year instrument record of global sea level rise. Since 2002, however, a pair of satellites known as the Gravity Recovery and Climate Experiment (GRACE) has observed changes in Earth's gravity from space. The gravity data are used to infer the redistribution of water mass across the surface and have provided the first direct measurements of changes in the mass of large glaciers, ice sheets, groundwater, and even the global oceans. GRACE data have shown that Antarctica is losing mass overall (Shepherd and Wingham, 2007) and that Greenland mass loss accelerated between 2002 and 2008 (Witze, 2008). Between 2003 and 2007, GRACE data suggest that the mass-related contribution to globally averaged sea level rise was 0.8 mm per year (Willis et al., 2008). Although GRACE is relatively new and interpretation of its data continues to evolve, it shows great promise as a powerful tool for observing and understanding the mass component [sea level rise](#) caused by the input of freshwater. This is extremely important as much of the uncertainty in sea level projections arises from a lack of understanding of the response of glaciers and ice sheets to anthropogenic climate change.

Future projections

Despite the socioeconomic importance of sea level rise, projections of future sea level remain problematic and controversial (Rahmstorf et al., 2006; Pfeffer et al., 2008). Global climate models predict between 20 and



Sea Level Rise, Figure 2 Change in sea level between 2006 and 2007 from satellite observations (*top panel*). Also shown are the changes in sea surface height caused by thermal expansion in the upper 750 m from the same period based on ocean temperature observations (*bottom panel*) (Updated from Willis et al., 2004).

60 cm of rise over the next 100 years (IPCC, 2007b) due to thermal expansion and simple melting process from glaciers and ice sheets. However, paleoclimate evidence (Rohling et al., 2008; Siddall and Kaplan, 2008) suggests that ice sheets can respond to climate change much more rapidly than is allowed for in the current generation of climate models. This is somewhat alarming given the several meters of sea level equivalent of freshwater contained in Greenland and the West Antarctica, the two ice sheets thought to be most responsive to climate change. Although complete loss of these ice sheets within the next 100 years seems highly unlikely, the range of possible loss rates remains a matter of scientific debate.

Summary

Sea level continues to rise at a rate that is large compared to any average rate over the past few 1,000 years. Furthermore, tide gauge and satellite data have shown that sea level rise has accelerated from an average rate of about 1.8 mm per year over the past 100 years to about 3 mm per year between 1992 and 2008. Although the instrumental record of thermal expansion in the upper ocean is much shorter, it accounted for about 1/3 of sea level rise between 1992 and 2008 and about one-fourth of total sea level rise between 1960 and 2000. Satellite observations of sea level, ocean mass, and the wasting of glaciers and ice sheets have revolutionized understanding of sea level rise and its causes, but projections of sea level rise remain highly uncertain.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Bindoff, N. L., Willebrand, J., Artale, V., Cazenave, A., Gregory, J., Gulev, S., Hanawa, K., Le Quéré, C., Levitus, S., Nojiri, Y., Shum, C. K., Talley, L. D., and Unnikrishnan, A., 2007. Observations: oceanic climate change and sea level. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK/New York: Cambridge University Press.
- Cazenave, A., and Nerem, R. S., 2004. Present-day sea level change: observations and causes. *Reviews of Geophysics*, **42**, RG3001, doi:10.1029/2003RG000139.
- Church, J. A., and White, N. J., 2006. A 20th century acceleration in global sea-level rise. *Geophysical Research Letters*, **33**, L01602, doi:10.1029/2005GL024826.
- Church, J. A., White, N. J., Aarup, T., Wilson, W. S., Woodworth, P. L., Domingues, C. M., Hunter, J. R., and Lambeck, K., 2008. Understanding global sea levels: past, present and future. *Sustainability Science*, **3**(1), 9–22, doi:10.1007/s11625-008-0042-4. Special Feature: Original Article.
- Domingues, C. M., Church, J. A., White, N. J., Gleckler, P. J., Wijffels, S. E., Barker, P. M., and Dunn, J. R., 2008. Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*, **453**, 1090–1093, doi:10.1038/nature07080.
- IPCC, 2007a. Summary for policymakers. In Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J., and Hanson, C. E. (eds.), *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press, pp. 7–22.
- IPCC, 2007b. Global climate projections. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK/New York: Cambridge University Press, 996 pp.
- Jansen, E., Overpeck, J., Briffa, K. R., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W. R., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R., and Zhang, D., 2007. Palaeoclimate. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK/New York: Cambridge University Press.
- Mitrovica, J. X., Tamiseia, M. E., Davis, J. L., and Milne, G. A., 2001. Recent mass balance of polar ice sheets inferred from pattern of global sea level change. *Nature*, **409**, 1026–1029.
- Pfeffer, W. T., Harper, J. T., and O'Neel, S., 2008. Kinematic constraints on glacier contributions to 21st-century sea-level rise. *Science*, **321**, 1340–1343, doi:10.1126/science.1159099.
- Rohling, E. J., Grant, K., Hemleben, C., Siddall, M., Hoogakker, B. A. A., Bolshaw, M., and Kucera, M., 2008. High rates of sea-level rise during the last interglacial period. *Nature Geoscience*, **1**, 38–42, doi:10.1038/ngeo.2007.28.
- Shepherd, A., and Wingham, D., 2007. Recent sea-level contributions of the Antarctic and Greenland ice sheets. *Science*, **16**, 1529–1532, doi:10.1126/science.1136776.
- Siddall, M., and Kaplan, M. R., 2008. Climate science: a tale of two ice sheets. *Nature Geoscience*, **1**, 570–572, doi:10.1038/ngeo286.
- Willis, J. K., Roemmich, D., and Cornuelle, B., 2004. Interannual variability in upper ocean heat content, temperature, and thermocline expansion on global scales. *Journal of Geophysical Research*, **109**, C12036, doi:10.1029/2003JC002260.
- Willis, J. K., Chambers, D. P., and Nerem, R. S., 2008. Assessing the globally averaged sea level budget on seasonal to interannual timescales. *Journal of Geophysical Research*, **113**, C06015, doi:10.1029/2007JC004517.
- Witze, A., 2008. Climate change: losing Greenland. *Nature*, **452**, 798–802.

Cross-references

[Climate Data Records](#)
[Climate Monitoring and Prediction](#)
[Cryosphere, Climate Change Effects](#)
[Earth Radiation Budget, Top-of-Atmosphere Radiation](#)
[Ice Sheets and Ice Volume](#)
[Ocean Surface Topography](#)
[Water and Energy Cycles](#)

SEA SURFACE SALINITY

Gary Lagerloef
ESR, Seattle, WA, USA

Introduction

Surface salinity is an ocean state variable that, along with temperature, controls the density of seawater and influences surface circulation and formation of dense surface waters in the higher latitudes which sink into the deep ocean and help drive the thermohaline circulation. The growing scientific need for global measurements stimulated the development of salinity-observing satellite sensors that have been launched in recent years. These are the European Soil Moisture Ocean Salinity (SMOS) and the US-Argentine Aquarius/SAC-D missions. Salinity remote sensing is possible because the dielectric properties of seawater which depend on salinity also affect the surface emission at certain microwave frequencies. Experimental heritage extends more than 35 years in the past, including laboratory studies, airborne sensors, and one instrument flown briefly in space on Skylab. Requirements for very low noise microwave radiometers and large antenna structures historically have limited the advance of satellite systems. New technologies in the current missions have addressed these issues. Science needs, primarily for climate studies, dictate a resolution requirement of approximately 100–200 km spatial grid, observed monthly, with approximately 0.1–0.2 error on the Practical Salinity Scale (or 1 part in 10,000). These demand very precise radiometers and that several ancillary errors be accurately corrected. Measurements are made in the 1.413 GHz astronomical hydrogen absorption band to avoid radio interference.

Definition and theory

How salinity is defined and measured

Salinity represents the concentration of dissolved inorganic salts in seawater (grams salt per kilogram seawater, or parts per 1,000, and historically given by the symbol ‰). Oceanographers have developed modern techniques based on the electrical conductivity of seawater which permit accurate measurement by use of automated electronic in situ sensors. Salinity is derived from conductivity, temperature, and pressure with an international standard set of empirical equations known as the *Practical Salinity Scale*, established in 1978 (PSS-78) which is much easier to standardize and more precise than previous chemical methods and which numerically represents grams per kilogram. Accordingly, the modern literature often quotes salinity measurements in practical salinity units (psu) or refers to PSS-78. Salinity ranges from near zero adjacent to the mouths of major rivers to more than 40 in the Red Sea. Aside from such

extremes, open ocean surface values away from coastlines generally fall between 32 and 37 (Fig. 1).

Historically, the global mean surface salinity field has been compiled from all available oceanographic observations dating back to the nineteenth century. Modern autonomous profiling buoys (Argo) are now providing near-real-time data from about 3,000 randomly distributed locations every 10 days. Nevertheless, a significant fraction of the 1° latitude longitude cells have no observations, requiring such maps as Fig. 1 to be interpolated and smoothed over scales of several 100 km. Sampling remains most sparse over large regions of the southern hemisphere. Remote sensing from satellites will be able to fill these gaps and monitor multiyear variations globally.

Remote sensing theory

Salinity remote sensing with microwave radiometry is likewise possible through the electrically conductive properties of seawater. A radiometric measurement of an emitting surface is given in terms of a *brightness temperature* (T_B), measured in degrees Kelvin (K). T_B is related to the true absolute surface temperature (T) through the emissivity coefficient (e):

$$T_B = eT$$

For seawater, e depends on the complex dielectric constant (ϵ), the viewing angle (Fresnel laws), and the surface roughness (due to wind waves). The complex dielectric constant is governed by the Debye equation, in simplified form:

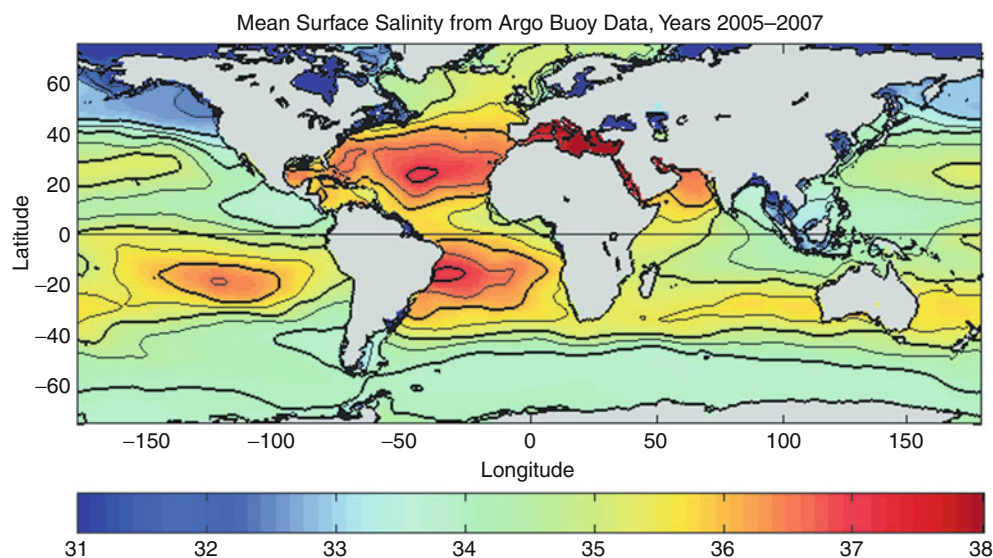
$$\epsilon = \epsilon_\infty + \frac{\epsilon_s(S, T) - \epsilon_\infty}{1 + i2\pi f\tau(S, T)} - \frac{iC(S, T)}{2\pi f\epsilon_0}$$

and includes electrical conductivity (C), the static dielectric constant (ϵ_s), and the relaxation time (τ) which are all sensitive to salinity and temperature (S, T). The equation also includes radio frequency (f) and terms for permittivity at infinite frequency (ϵ_∞) which may vary weakly with T and permittivity of free space (ϵ_0) which is a constant. $C(S, T)$ is derived from the *Practical Salinity Scale*. The static dielectric and time constants have been modeled by making laboratory measurements of ϵ at various frequencies, temperatures, and salinities and fitting ϵ_s and τ to polynomial expressions of (S, T) to match the ϵ data. Different models in the literature show similar variations with respect to (f, S, T).

Emissivity for the horizontal (H) and vertical (V) polarization state is related to ϵ by Fresnel reflection:

$$e_H = 1 - \left[\frac{\cos\theta - (\epsilon - \sin^2\theta)^{1/2}}{\cos\theta + (\epsilon - \sin^2\theta)^{1/2}} \right]^2;$$

$$e_V = 1 - \left[\frac{\epsilon\cos\theta - (\epsilon - \sin^2\theta)^{1/2}}{\epsilon\cos\theta + (\epsilon - \sin^2\theta)^{1/2}} \right]^2$$



Sea Surface Salinity, Figure 1 Contour map of the mean global surface salinity field (contour interval 0.5) averaged over the years 2005–2007, interpolated from ocean buoy data (Argo), and supplemented with data from the World Ocean Atlas, 2005 (U.S. Department of Commerce, NOAA, National Oceanographic Data Center). Surface salinity ranges from about 32 to 37 psu. Highest values are in the high-evaporation regions of the subtropics, especially in the Atlantic Ocean, and the Mediterranean Sea. Lowest salinity values coincide with regions with the highest rainfall.

where θ is the vertical incidence angle from which the radiometer views the surface and $e_H = e_V$ when $\theta = 0$. The above set of equations provides a physically based model function relating T_B to surface S , T , θ , and H or V polarization state for smooth water (no wind roughness). This can be inverted to retrieve salinity from radiometric T_B measurements provided the remaining parameters are known. The microwave optical depth is such that the measured emission originates in the top 1 cm of the ocean, approximately.

The rate at which T_B varies with salinity is sensitive to microwave frequency, achieving levels practical for salinity remote sensing at frequencies less than about 3 GHz. Considerations for selecting a measurement radio frequency include salinity sensitivity, requisite antenna size (see below), and radio interference from other (mostly man-made) sources. A compromise of these factors, dominated by the interference issue, dictates a choice of about a 27 MHz-wide frequency band centered at 1.413 GHz, which is the hydrogen absorption band protected by international treaty for radio astronomy research. This falls within a frequency range known as L-band. Atmospheric clouds have a negligible effect, allowing observations in all weather except possibly heavy rain. Accompanying illustrations are based on applying $f = 1.413$ GHz in the Debye equation and using a model that included laboratory dielectric constant measurements at the nearby frequency 1.43 GHz. Figure 2 shows how T_B varies as a function of surface T and S for V and H polarization and incidence angle of 45.6° , which is one of the three incidence angles viewed by Aquarius/SAC-D (along with 28.7° and 37.8°). Note that for either polarization, it is

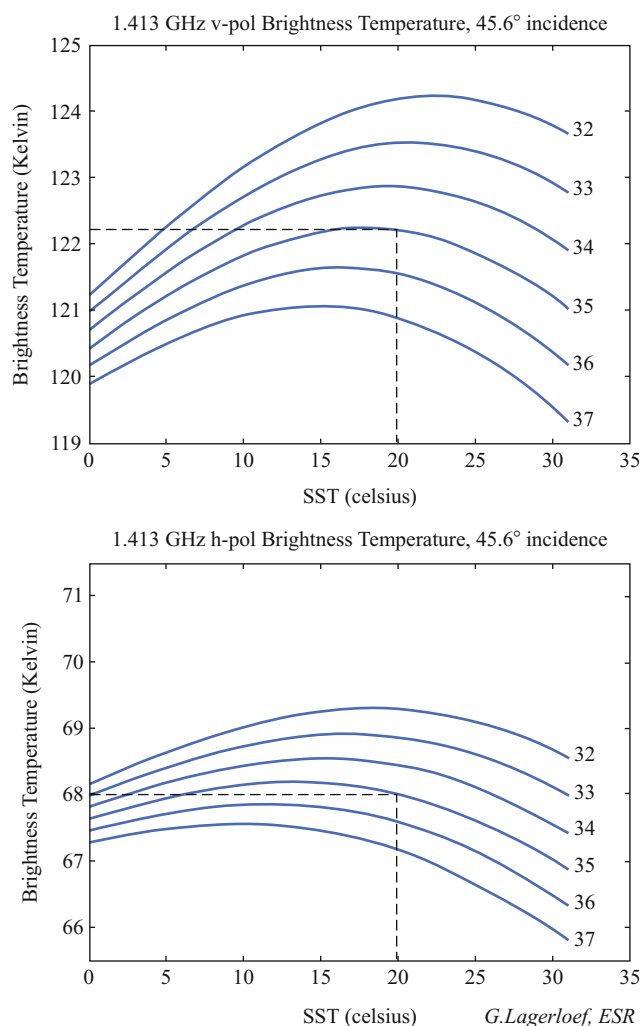
possible to retrieve salinity if both T and T_B are precisely known at either polarization (dashed lines). In practice, both polarizations will be measured and combined to improve the retrieval accuracy and possibly to determine other parameters such as surface roughness. (Note that at $\theta = 0$, V and H polarization brightness temperatures are identical.) Other features of this model function and their influence on measurement accuracy are discussed in a later section on resolution and error sources.

Antennas

Unusually large radiometer antennas will be required to be deployed on satellites to measure salinity. Radiometer antenna beamwidth varies inversely with both antenna aperture and radio frequency. 1.413 GHz is a significantly lower frequency than that found on conventional satellite microwave radiometers, and large antenna structures are necessary to avoid excessive beamwidth and accordingly large footprint size. For example, a 50 km footprint requires about a 6 m aperture antenna, whereas conventional radiometer antennas are around 1–2 m. To decrease the footprint by a factor of two requires doubling the antenna aperture, whereas reducing the aperture increases the footprint proportionately. Various filled and thinned array technologies for large antennas have now reached a development stage where application to salinity remote sensing is feasible.

History of salinity remote sensing

The first experiment to measure surface salinity from space took place on the NASA Skylab mission during the fall and winter 1973–1974, when a 1.413 GHz



Sea Surface Salinity, Figure 2 Brightness temperature (T_B) as a function of S , T , and polarization (V and H) for typical ocean surface conditions and incidence angle of 45.6° (as an example from one to three Aquarius/SAC-D satellite viewing angles). Salinity can be determined from either polarization if both T_B and T are known (*dashed lines*) (calculations are based on formulas in Klein and Swift, 1977).

microwave radiometer with a 1 m antenna collected intermittent data. A weak correlation was found between the sensor data and surface salinity, after correcting for other influences. There was no “ground truth” other than standard surface charts, and many of the ambient corrections were not as well modeled then as they could be today. Research leading up to the Skylab experiment began with several efforts during the late 1940s and early 1950s to measure the complex dielectric constant of saline solutions for various salinities, temperatures, and microwave frequencies. These relationships provide the physical basis for microwave remote sensing of the ocean as described above.

The first airborne salinity measurements were demonstrated in the Mississippi River outflow and published in 1970. This led to renewed efforts during the 1970s to refine the dielectric constants and governing equations. Meanwhile, a series of airborne experiments in the 1970s mapped coastal salinity patterns in the Chesapeake and Savannah River plumes and freshwater sources along the Puerto Rico shoreline. In the early 1980s, a satellite concept was suggested that might achieve an ideal precision of about 0.25 and spatial resolution of about 100 km. At that time, space agencies were establishing the oceanic processes remote sensing programs around missions and sensors for measuring surface dynamic topography, wind stress, ocean color, surface temperature, and sea ice. For various technical reasons, salinity remote sensing was only marginally feasible from satellite and other types of measurements took precedence.

Interest in salinity remote sensing revived in the late 1980s with the development of a 1.4 GHz airborne Electrically Scanning Thinned Array Radiometer (ESTAR) designed primarily for soil moisture measurements. ESTAR imaging is done electronically with no moving antenna parts, thus making large antenna structures more feasible. The airborne version was developed as an engineering prototype and to provide the proof-of-concept that aperture synthesis can be extrapolated to a satellite design. The initial experiment to collect ocean data with this sensor consisted of a flight across the Gulf Stream in 1991 near Cape Hatteras. The change from 36 psu in the offshore waters to < 32 psu near shore was measured, along with several frontal features visible in the satellite surface temperature image from the same day. This Gulf Stream transect demonstrated that small salinity variations typical of the open ocean can be detected as well as the strong salinity gradients in the coastal and estuary settings demonstrated previously.

By the mid-1990s, a new airborne Scanning Low Frequency Microwave Radiometer (SLFMR) was developed to measure salinity from light aircraft and has been extensively used by NOAA and the US Navy to survey coastal and estuary waters on the US east coast and Florida. A version of this sensor is now being used in Australia, and a second-generation model is presently used for research by the US Navy.

In 1999, a satellite project was approved by the European Space Agency for the measurement of Soil Moisture and Ocean Salinity (SMOS), which was launched in November 2009. The SMOS mission design emphasizes soil moisture measurement requirements, which is done at the same microwave frequency for many of the same reasons as salinity. The T_B dynamic range is about 70–80 K for varying soil moisture conditions, and the precision requirement is, therefore, much less rigid than for salinity. SMOS employs a 6 m aperture two-dimensional phased array antenna system that yields 40–90 km resolution across the measurement swath. For the ocean, SMOS data are averaged over 200 km and 10–30 day scales to reduce measurement noise.

The Aquarius/SAC-D is a joint mission of the USA (NASA) and Argentina (CONAE) and was launched in June 2011. The Aquarius/SAC-D design puts primary emphasis on ocean salinity rather than soil moisture, with the focus on optimizing salinity accuracy with a very precisely calibrated microwave radiometer and key ancillary measurements for addressing the most significant error sources. With this advanced radiometer system and a 2.5 m aperture antenna, Aquarius/SAC-D will yield 0.2 psu accuracy on 150 km scale, monthly.

Requirements for observing salinity from satellite Scientific issues

Three broad scientific themes have been identified for a satellite salinity remote sensing program. These themes relate directly to the international climate research and global environmental observing program goals.

Improving seasonal to interannual climate predictions

This focuses primarily on El Niño forecasting and involves the effective use of surface salinity data (1) to initialize and improve the coupled climate forecast models and (2) to study and model the role of freshwater flux in the formation and maintenance of barrier layers and mixed layer heat budgets in the tropics. Climate prediction models in which satellite altimeter sea level data are assimilated must be adjusted for steric height (sea level change due to ocean density) caused by the variations in upper layer salinity. If not, the adjustment for model heat content is incorrect and the prediction skill is degraded. Barrier layer formation occurs when excessive rainfall creates a salinity-stratified layer between a well-mixed surface layer and the thermocline, which effectively isolates the deeper thermocline from exchanging heat with the atmosphere with consequences on the air-sea coupling processes that, for example, govern El Niño dynamics.

Improving ocean rainfall estimates and global hydrologic budgets

Precipitation over the ocean is still poorly quantified and relates to both the hydrologic budget and to latent heating of the overlying atmosphere. Using the ocean as a rain gauge is feasible with precise surface salinity observations coupled with ocean surface current velocity data and mixed layer modeling. Such calculations will reduce uncertainties in the surface freshwater flux on climate time scales and will complement satellite precipitation and evaporation observations to improve estimates of the global water and energy cycles.

Monitoring large-scale salinity events and thermohaline convection

Studying interannual surface salinity variations in the subpolar regions, particularly the North Atlantic and Southern Oceans, is essential to long time scale climate prediction and modeling. These variations influence the rate of oceanic convection and poleward heat transport (thermohaline circulation) which are known to have been

coupled to extreme global climate changes in the geologic record. Outside of the polar regions, salinity signals are stronger in the coastal ocean and marginal seas than in the open ocean in general, but large footprint size will limit near-shore applications of the data. Many of the larger marginal seas which have strong salinity signals might be adequately resolved nonetheless, such as the East China Sea, Bay of Bengal, Gulf of Mexico, Coral Sea and Gulf of Papua, and the Mediterranean Sea.

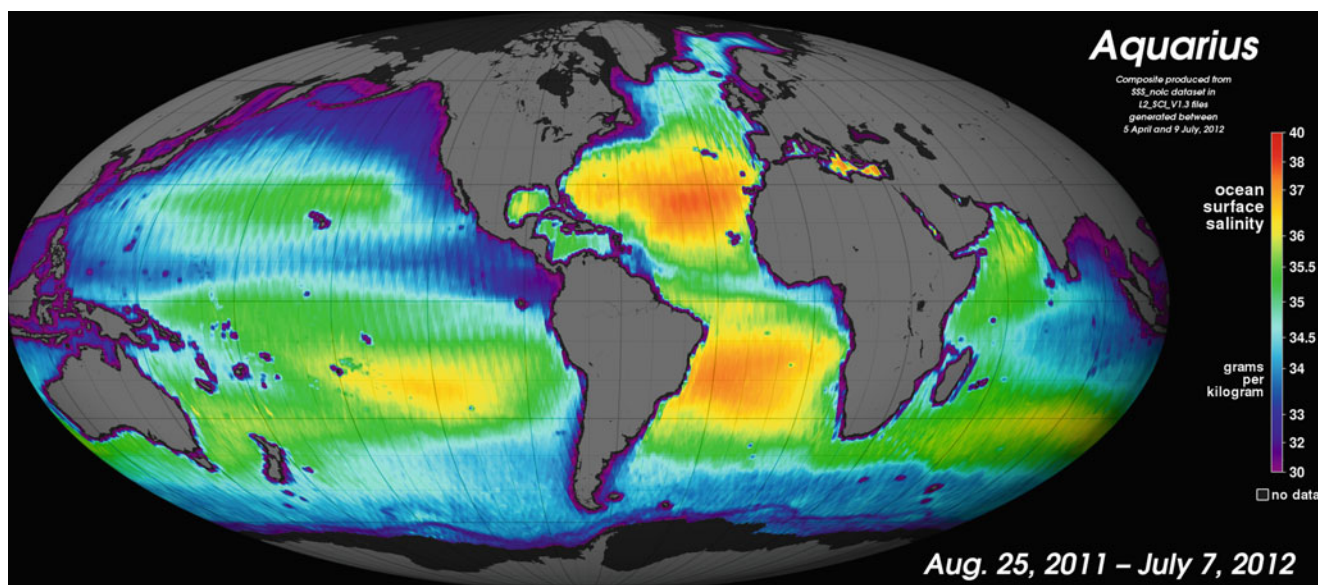
Science requirements

The following four topics derive from these science themes. Preliminary accuracy, spatial, and temporal resolution requirements have been suggested as the minimum to derive scientific benefit from satellite observations: (1) barrier layer effects on tropical Pacific heat flux 0.2 (PSS-78), 100 km, 30 days; (2) steric adjustment of heat storage from sea level 0.2, 200 km, 7 days; (3) North Atlantic thermohaline circulation 0.1, 100 km, 30 days; and (4) surface freshwater flux balance 0.1, 300 km, 30 days. Thermohaline circulation and convection in the subpolar seas has the most demanding requirement and is the most technically challenging because of the reduced T_B /salinity ratio at low seawater temperatures (see below). This can serve as a prime satellite mission requirement, allowing for the others to be met by reduced mission requirements as appropriate. Aquarius/SAC-D is a pathfinder mission capable of meeting the majority of these requirements with a grid scale of 150 km and salinity error less than 0.2 psu observed monthly, as noted above. Figure 3 shows the global distribution of sea surface salinity derived from Aquarius measurements for the period August 25, 2011, to July 7, 2012. The near-latitudinal striping is an artifact of the sampling.

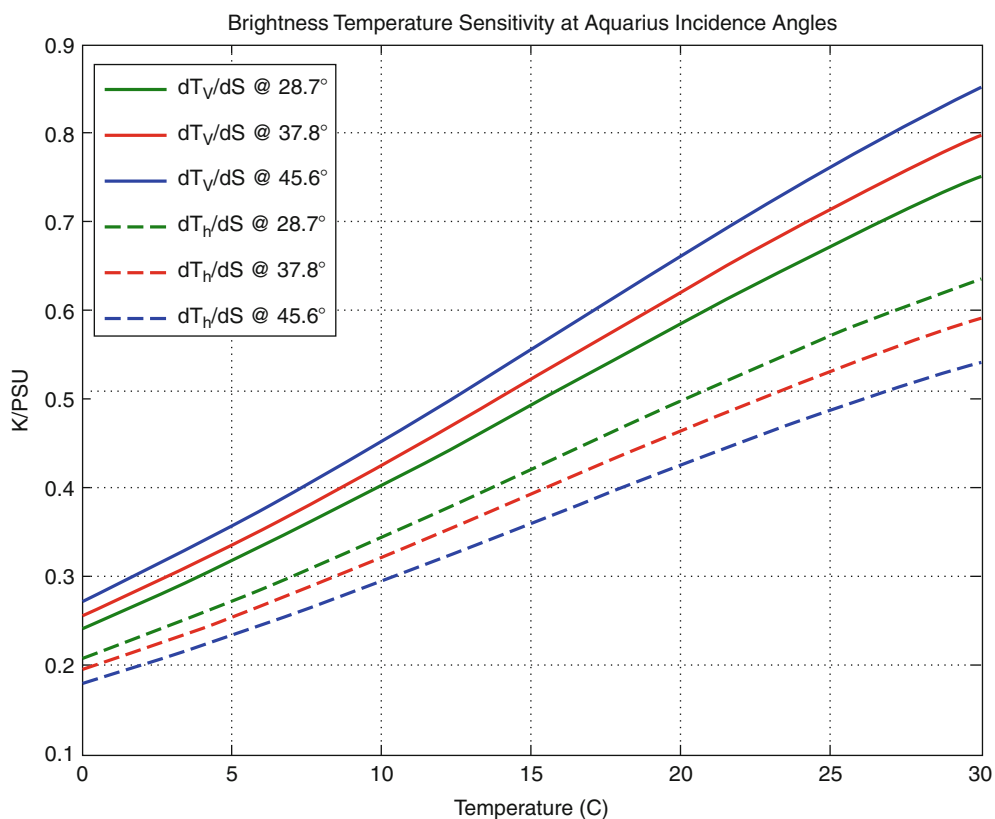
Resolution and error sources

Model function

From Fig. 2, it is evident that the dynamic range of T_B is about 4 K over the range of typical open ocean surface salinity and temperature conditions. At a given temperature, T_B decreases as salinity increases, whereas the tendency with respect to temperature changes sign. It is also evident that salinity contours are spread farther apart for V polarization than for H, and therefore, V is slightly more sensitive to salinity changes. The sensitivity also increases with temperature for each polarization. These properties are shown more clearly in Fig. 4. The differential of T_B with respect to salinity (absolute value) ranges from about 0.2 to 0.8 K per salinity unit. Corrected T_B will need to be measured to 0.02–0.08 K precision to achieve 0.1 psu salinity resolution. The difference in sensitivity between polarizations also increases with incidence angle. The sensitivity is strongly affected by temperature, being largest at the highest temperatures and yielding better measurement precision in warm versus cold ocean conditions. Random error can be reduced by temporal and spatial averaging. The degraded measurement precision in higher latitudes



Sea Surface Salinity, Figure 3 Global sea surface salinity derived from the measurements of the Aquarius instrument. Image provided by NASA (<http://aquarius.nasa.gov/gallery-science.html>).



Sea Surface Salinity, Figure 4 The rate that T_B varies with salinity change (sensitivity) depends on temperature, polarization, and incidence angle. The curves are shown here for the three Aquarius/SAC-D satellite mission incidence angles.

will be partly offset by averaging with the greater sampling frequency from a polar-orbiting satellite.

The T_B variation with respect to temperature falls generally between $\pm 0.15 \text{ K } ^\circ\text{C}^{-1}$ and near zero over a broad S and T range. Knowledge of the surface temperature to within a few tenths $^\circ\text{C}$ will be adequate to correct T_B for temperature effects and can be obtained using data from other satellite systems. The optical depth for this microwave frequency in seawater is about 1–2 cm, and the remotely sensed measurement depends on the T and S in that surface layer thickness. T_{BS} for the H and V polarizations have large variations with incidence angle and space craft attitude will need to be monitored very precisely. The accuracy in the proximity to land and ice boundaries will also be degraded, because these surfaces have much higher brightness temperatures than seawater. The best accuracy will be achieved over open water at distances of at least three footprint diameters from land or ice boundaries.

Other errors

Several other error sources will bias T_B measurements and must be either corrected or avoided. These include ionosphere and atmosphere effects, cosmic and galactic background radiation, surface roughness from winds, sun glint, solar flux, and rain effects. Cosmic background and lower atmospheric adsorption are nearly constant biases easily corrected. An additional correction will be needed when the reflected radiation from the galactic core is in the field of view. The ionosphere and surface winds (roughness) have wide spatial and temporal variations and require ancillary data and careful treatment to avoid T_B errors of several K.

The magnitude of the wind roughness correction varies with incidence angle and polarization and ranges between 0.1 and $0.4 \text{ K}/(\text{ms}^{-1})$. It will be the most difficult to correct for accurately and is consequently the largest uncertainty in the error budget. Sea state conditions change significantly within the few hours that may elapse until ancillary measurements are obtained from another satellite. Simultaneous wind roughness measurement can be made with an onboard radar backscatter sensor, such as is included on the Aquarius/SAC-D mission. This provides the most accurate correction possible by using a direct relationship between the radar backscatter and the T_B roughness effect, rather than rely on wind or sea state information from other sources.

The ionosphere affects the measurement through attenuation and through Faraday rotation of the H and V polarized signal. There is no Faraday effect when viewing at nadir ($\theta = 0$) because H and V emissivities are identical, whereas off-nadir corrections are needed to preserve the polarization signal. Correction data can be obtained from ionosphere models and analyses but may be limited by unpredictable short-term ionosphere variations. Onboard correction techniques have been developed that require fully polarimetric radiometer measurements from which the Faraday rotation may be

derived. Sun-synchronous orbits can be selected that minimize the daytime peak in ionosphere activity as well as solar effects.

Microwave attenuation by rain depends on rain rate and the thickness of the rain layer in the atmosphere. The effect is small at the microwave frequency used, but for the required accuracy, either the effect must be modeled and corrected with ancillary data or the contaminated data discarded. For the accumulation of all the errors described here, it is anticipated that the root sum square salinity error will be reduced to less than 0.2 with adequate radiometer engineering, correction models, onboard measurements, ancillary data, and spatiotemporal filtering with methods now in development.

Tables 1, 2 and 3 shows the accumulation of all these factors in an error budget analysis for Aquarius/SAC-D. There is an allocation for each error term, as well as a current best estimate (CBE). These are accumulated as a root sum square (RSS). The retrieval sensitivity as it varies with temperature is used to compute the mean salinity error by latitude band averaged over a month. The root mean square (RMS) of these yields the global monthly error. The CBE is less than the allocation; there is some RSS margin for additional error growth while still meeting the target of 0.2 psu global RMS error.

Sea Surface Salinity, Table 1 An error budget analysis for the Aquarius/SAC-D satellite

Error sources	3 beam RMS	
	Allocation	CBE
Radiometer	0.15	0.09
Antenna	0.08	0.01
System pointing	0.05	0.02
Roughness	0.28	0.20
Solar	0.05	0.02
Galactic	0.05	0.004
Rain (total liquid water)	0.02	0.01
Ionosphere	0.06	0.043
Atmosphere – other	0.05	0.02
SST	0.10	0.07
Antenna gain near land & ice	0.10	0.10
Model function	0.08	0.07

Sea Surface Salinity, Table 2 An error budget analysis for the Aquarius/SAC-D satellite

Brightness temperature error per observation	Baseline mission	
	Allocation	CBE
Total RSS (K)	0.38	0.27
Margin RSS (K)	0.27	

Sea Surface Salinity, Table 3 An error budget analysis for the Aquarius/SAC-D satellite

Latitude range	Mean sensitivity (dTv/dS)	Mean # samples in 28 days	Baseline mission monthly salinity error (psu)	
			Allocation	CBE
0–10	0.756	10.9	0.15	0.11
11–20	0.731	11.3	0.16	0.11
21–30	0.671	12.1	0.16	0.12
31–40	0.567	13.5	0.18	0.13
41–50	0.455	15.9	0.21	0.15
51–60	0.357	20.3	0.24	0.17
61–70	0.271	30.2	0.26	0.18
Global RMS (psu)			0.20	0.14
Margin RSS (psu)			0.14	

Summary

The technology to measure ocean surface salinity from space has advanced significantly during the past decade. The basic physical principles are well understood. New technologies for radiometer accuracy, large antenna systems, and various ambient corrections have overcome obstacles that inhibited salinity satellite remote sensing in the past. The Aquarius/SAC-D and SMOS satellites provide the unprecedented capability to map the global open ocean salinity field with the necessary resolution and accuracy, which is currently being assessed, to significantly advance the scientific understanding of the links between the ocean circulation, the global water cycle, and the climate.

Bibliography

- Blume, H. C., and Kendall, B. M., 1982. Passive microwave measurements of temperature and salinity in coastal zones. *GE: IEEE Transactions on Geoscience and Remote Sensing*, 394–404.
- Blume, H.-J. C., Kendall, B. M., and Fedors, J. C., 1978. Measurement of ocean temperature and salinity via microwave radiometry. *Boundary-Layer Meteorology*, 13, 295–380.
- Blume, H.-J. C., Kendall, B. M., and Fedors, J. C., 1981. Multifrequency radiometer detection of submarine freshwater sources along the Puerto Rican coastline. *Journal of Geophysical Research*, 86, 5283–5291.
- Broecker, W. S., 1991. The great ocean conveyor. *Oceanography*, 4, 79–89.
- Delcroix, T., and Henin, C., 1991. Seasonal and interannual variations of sea surface salinity in the tropical Pacific Ocean. *Journal of Geophysical Research*, 96, 22135–22150.
- Delworth, T., Manabe, S., and Stouffer, R. J., 1993. Interdecadal variations of the thermohaline circulation in a coupled ocean–atmosphere model. *Journal of Climate*, 6, 1993–2011.
- Dickson, R. R., Meincke, R., Malmberg, S.-A., and Lee, J. J., 1988. The ‘Great Salinity Anomaly’ in the northern North Atlantic, 1968–1982. *Progress in Oceanography*, 20, 103–151.
- Droppelman, J. D., Mennella, R. A., and Evans, D. E., 1970. An airborne measurement of the salinity variations of the Mississippi river outflow. *Journal of Geophysical Research*, 75, 5909–5913.
- Font, J., Lagerloef, G. S. E., Le Vine, D. M., Camps, A., and Zanife, O. Z., 2004. The determination of surface salinity with the

European SMOS Space mission. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 2196–2205.

- Kendall, B. M., and Blanton, J. O., 1981. Microwave radiometer measurement of tidally induced salinity changes off the Georgia coast. *Journal of Geophysical Research*, 86, 6435–6441.
- Klein, L. A., and Swift, C. T., 1977. An improved model for the dielectric constant of sea water at microwave frequencies. *IEEE Transactions on Antennas and Propagation*, AP-25(1), 104–111.
- Lagerloef, G. S. E., 2000. Recent progress toward satellite measurements of the global sea surface salinity field. In Halpern, D. (ed.), *Satellites, Oceanography and Society*. Elsevier Oceanography Series, 63, 367 pp.
- Lagerloef, G., Swift, C., and LeVine, D., 1995. Sea surface salinity: the next remote sensing challenge. *Oceanography*, 8, 44–50.
- Lagerloef, G. S. E., Colomb, F. R., Le Vine, D. M., Wentz, F. J., Yueh, S. H., Ruf, C. S., Lilly, J., Gunn, J., Chao, Y., de Charon, A., Feldman, G., and Swift, C. T., 2008. The Aquarius/SAC-D mission. *Oceanography*, 21, 68–81.
- Le Vine, D. M., Kao, M., Tanner, A. B., Swift, C. T., and Griffiths, A., 1990. Initial results in the development of a synthetic aperture radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4), 614–619.
- Le Vine, D. M., Kao, M., Garvine, R. W., and Sanders, T., 1998. Remote sensing of ocean salinity: results from the Delaware current experiment. *Journal of Atmosphere and Ocean Technical*, 15, 1478–1484.
- Le Vine, D. M., Lagerloef, G. S. E., Colomb, R., Yueh, S., and Pellerano, F., 2007. Aquarius: an instrument to monitor sea surface salinity from space. *IEEE Transactions on Geoscience and Remote Sensing*, 45, 2040–2050.
- Lerner, R. M., and Hollinger, J. P., 1977. Analysis of 1.4 GHz radiometric measurements from Skylab. *Remote Sensing Environment*, 6, 251–269.
- Lewis, E. L., 1980. The practical salinity scale 1978 (PSS-78) and its antecedents. *IEEE Journal on Oceanic Engineering*, OE-5, -8.
- Meissner, T., and Wentz, F. J., 2003. The complex dielectric constant of pure and sea water from microwave satellite observations. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1836–1849.
- Miller, J., Goodberlet, M., and Zaitzeff, J., 1998. Airborne salinity mapper makes debut in coastal zone. *EOS Transactions, American Geophysical Union*, 79, 173 & 176–177.
- Reynolds, R., Ji, M., and Leetmaa, A., 1998. Use of salinity to improve ocean modeling. *Physics and Chemistry of the Earth*, 23, 545–555.
- Swift, C. T., and McIntosh, R. E., 1983. Considerations for microwave remote sensing of ocean-surface salinity. *IEEE Transactions on Geoscience and Remote Sensing*, GE-21, 480–491.
- UNESCO, 1981. The practical salinity scale 1978 and the international equation of state of seawater 1980. *Technical Papers in Marine Science*, 36, 25 pp.
- Wilson, W. J., Yueh, S. H., Dinardo, S. J., and Li, F. K., 2004. High-stability L-band radiometer measurements of saltwater. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1829–1835.
- Yueh, S. H., West, R., Wilson, W. J., Li, F. K., Njoku, E. G., and Rahmat-Samii, Y., 2001. Sources and feasibility for microwave remote sensing of ocean surface salinity. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1049–1060.

Cross-references

Climate Data Records
Climate Monitoring and Prediction
Cloud Liquid Water

[Cryosphere, Climate Change Feedbacks](#)
[Microwave Dielectric Properties of Materials](#)
[Microwave Horn Antennas](#)
[Microwave Radiometers](#)
[Microwave Radiometers, Conventional](#)
[Microwave Radiometers, Correlation](#)
[Microwave Radiometers, Interferometers](#)
[Microwave Surface Scattering and Emission](#)
[Ocean, Measurements and Applications](#)
[Ocean Modeling and Data Assimilation](#)
[Reflector Antennas](#)
[Soil Moisture](#)
[Surface Truth](#)
[Water and Energy Cycles](#)

SEA SURFACE TEMPERATURE

Peter J. Minnett
 Rosenstiel School of Marine and Atmospheric Science,
 University of Miami, Miami, FL, USA

Definition

Historically, [sea surface temperature \(SST\)](#) has been measured by mercury-in-glass thermometers in buckets containing seawater collected by lowering them from ships and, in recent decades, thermistors on buoys or in the engine cooling water intakes of ships. These thermometers measure a temperature at depths of tens of centimeters to a few meters, a temperature often referred to as the “bulk” SST. Recently it has become accepted that it is better to include the depth of the measurement as vertical gradients can exist in the top few meters of the water column. The remotely sensed measurements of SST from satellite infrared radiometers are of a temperature very close to the sea surface, often referred to as the skin temperature. The skin temperature is close to the temperature of the ocean that is in contact with the atmosphere and, as such, is an important parameter that controls the heat and gas transfers between the ocean and atmosphere.

Near-surface temperature gradients

There are physical processes that introduce temperature differences in the uppermost meters of the ocean: the skin effect and diurnal, or daily, heating and cooling. The heat loss to the atmosphere cools the surface by typically between 0.1 and 0.5 K. Diurnal heating can raise afternoon temperatures by several degrees, although ~0.5 K is more typical. These changes are significant given that over much of the world’s oceans the seasonal signal in SST is less than 5 K. Similarly, for the major interannual weather anomalies, such as El Niño, the amplitudes of the associated SST signals are only a few degrees.

Diurnal heating

The ocean absorbs sunlight in the uppermost meters to tens of meters, depending on the water clarity, and as

a result becomes warmer (e.g., Gentemann and Minnett, 2008; Gentemann et al., 2008). When the sea is calm, the heating remains close to the surface, whereas in windy conditions, the turbulence in the upper ocean mixes the heat through the so-called mixed layer, which is typically many tens to a hundred meters deep or more. Thus, in calm conditions, the heat absorption leads to a noticeable increase in SST, but in windy situations, the diurnal SST change can be very small. The diurnal heating is often erased at night as the ocean gives up the heat stored during the day to the atmosphere. The summation over time of the net daily change in SST and in upper ocean heat content gives the seasonal cycles of these variables, increasing in the spring and summer and decreasing in the autumn and winter. In nearly all situations, the ocean is warmer than the overlying surface layer of the atmosphere.

Skin effect

Within the ocean and lower atmosphere, the vertical transport of heat is accomplished by turbulence, but the heat flow from the ocean to the atmosphere occurs through molecular conduction. This requires a temperature gradient to exist as heat is conducted from a warmer to a cooler region. For most situations, the ocean is warmer than the overlying surface layer of the atmosphere. Thus, through molecular conduction, the surface of the ocean, the skin layer, is generally cooler than the water just below, in the subskin layer (for an explanation of the nomenclature, see Donlon et al., (2007)). The temperature difference across the skin layer is dependent on several parameters including wind speed and approaches as constant value of about −0.14 K at high winds (Donlon et al., 2002; Minnett et al., 2011).

Remote sensing of Sea surface temperature

The remote sensing of SST from satellites is based on measuring the thermal emission of electromagnetic radiation from the sea surface, using instruments called radiometers that measure the radiative energy flux at the satellite height in distinct regions of the infrared and microwave parts of the electromagnetic spectrum. The spectrum of the natural emission of radiant energy $B_\lambda(T)$ from a perfectly emitting surface, a “black-body,” at temperature T K is given by the Planck equation:

$$B_\lambda(T) = 2hc^2\lambda^{-5} \left(e^{hc/(\lambda kT)} - 1 \right)^{-1}$$

where h is Planck’s constant, c is the speed of light in a vacuum, k is Boltzmann’s constant, and λ is the wavelength of the radiation. The sea surface, as all natural surfaces, is not a perfect black body, and the radiant emission is reduced by a factor called the emissivity that is wavelength and emission angle dependent, $\varepsilon(\lambda, \theta)$. The black body temperature corresponding to the radiative flux at the satellite height, the “brightness temperature,” is not equal to the skin SST because $\varepsilon < 1$ and the electromagnetic radiation is modified by absorption, emission,

and scattering as it propagates through the atmosphere. The spectrum of atmospheric transmissivity varies between zero and one, and satellite measurements of SST are made where the atmosphere is moderately transmissive. This limits measurements to two atmospheric transmission “windows” in the infrared, at wavelengths of 3.5–4.1 and 8–12 μm . In the infrared, ε of seawater is very high, 0.95–0.98, decreasing for emission angles away from normal. The derivation of SST to useful accuracy from infrared measurements requires correction of the atmospheric and surface effects. For microwaves the frequency range where SST measurements are most feasible is 6–10 GHz, but ε is quite small (0.3–0.5) and strongly dependent on the surface roughness.

Most satellite-derived SSTs come from instruments on polar-orbiting satellites. These orbit the Earth at about 800 km above the surface and take about 100 min for a complete revolution of the Earth, passing close to the poles. Satellite radiometers take measurements across a broad swath of the Earth by using rotating mirrors or push-broom detectors in the infrared and rotating antennas in the microwave. The area scanned in one rotation is incremented by the satellite orbital motion to generate a continuous image, generally centered on the subsatellite track. In the time it takes the satellite to complete an orbit, the Earth’s rotation moves the area seen from the satellite to the west so that the entire Earth’s surface is eventually imaged. Depending on the width of the swath, complete coverage can generally be achieved in 1–3 days.

An important prerequisite for the accurate retrieval of SST from both infrared and microwave radiometers is that the brightness temperature measurements be well calibrated. This requires careful characterization of the instruments prior to launch and, very importantly, accurate calibration of the sensors once in orbit. This is achieved by building into the radiometer calibration targets which have known emissivity and whose temperatures are accurately monitored. A natural calibration target often used in conjunction with the onboard target is deep space, away from the Sun, which has a brightness temperature of 2.7 K.

Infrared radiometers

A major limitation on the derivation of SST from infrared radiometers is the presence of clouds in the intervening atmosphere as they block the emission from the sea surface from reaching the satellite height. The clouds, of course, contribute their own emission to the radiometer measurement. Thus, any measurements containing clouds in the field of view of the radiometer must be identified and removed from the SST retrieval process (e.g., Saunders and Kriebel, 1988; Kilpatrick et al., 2001; Merchant et al., 2005). Because of the need for thorough screening of clouds in the images, all infrared radiometers designed for SST retrieval include channels responsive to scattered solar radiation (in the visible or near-infrared) to help identify clouds, at least during the sunlit part of each orbit.

Atmospheric water vapor absorbs a significant fraction of the sea surface emission in the infrared atmospheric windows, even in clear-sky conditions. The variability of water vapor requires an atmospheric correction algorithm to be applied for the actual, rather than mean, conditions at the time and place of the measurements. This can be accomplished by using measurements taken in distinct spectral intervals in the atmospheric transmission windows where the water vapor effect is different. These spectral intervals are defined by the characteristics of the radiometer and are called “bands” or “channels.” The idea is that the differences in the brightness temperatures measured in two or more channels at different wavelengths are related to the difference between the SST and the brightness temperature in one of them. Despite the complexity of the interactions between the infrared radiation and the molecules in the atmosphere, simple formulation can be used to derive SST to useful accuracy. One such formulation that is currently widely used is the nonlinear SST algorithm (Walton et al., 1998; May et al., 1998):

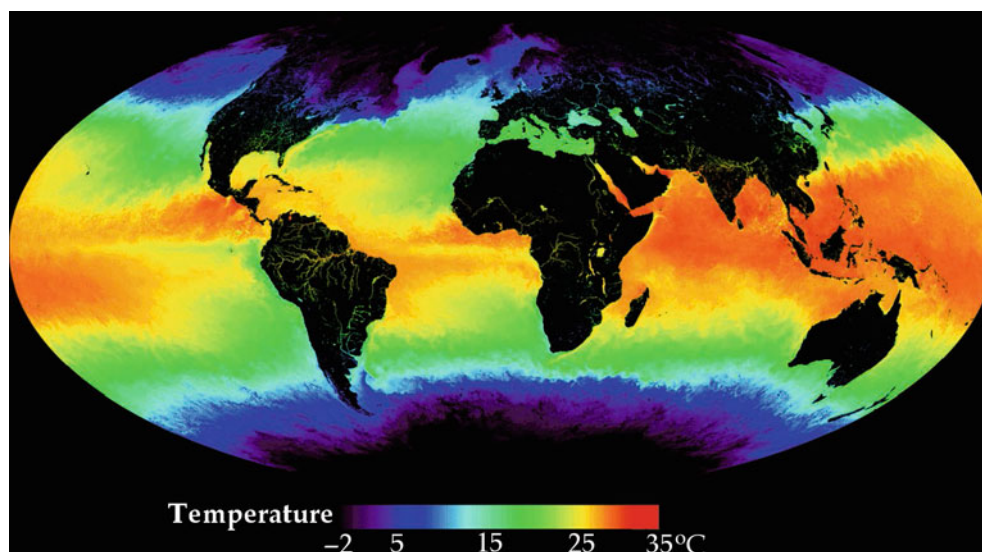
$$SST = a_0 + a_1 T_{11} + a_2 (T_{11} - T_{12}) SST_r + a_3 (T_{11} - T_{12}) (\sec(\theta) - 1)$$

where SST_r is a reference SST (or first-guess temperature) such as one based on climatology of the area and θ is the zenith angle to the satellite radiometer measured at the sea surface. The angular term accommodates the effects of increasing path length for measurements away from nadir and for the angular dependence of ε . The values of the coefficients a_i are usually determined either by regression analysis of coincident satellite and in situ measurements, mainly from buoys (Kilpatrick et al., 2001), or of simulated satellite measurements using radiative transfer modeling (Llewellyn-Jones et al., 1984; Merchant et al., 2004).

The measurements in the 3.7 μm atmospheric window include reflected and scattered sunlight during the daytime part of each orbit, and so these data are generally used for SST measurements only at night. When used in combination with measurements in the 10–12 μm range, in an algorithm similar to the NLSST, the derived SSTs are generally more accurate than those derived from the measurements in the 10–12 μm range alone.

With improvements in the fidelity of the representation of the atmospheric state in weather forecast models and analyses, in knowledge of the infrared spectroscopy of atmospheric gases, and in computer speed, alternative approaches to the atmospheric correction that make use of radiative transfer simulation of the measurements are being developed and applied (Merchant et al., 2009). These approaches, using optimal estimation techniques, offer potential improvements over the structural limitations of the simple NLSST formulation.

The longest time series of satellite-derived SST measurements has come from the Advanced Very High Resolution Radiometers (AVHRR (Cracknell, 1997)).



Sea Surface Temperature, Figure 1 The global sea surface temperature for May 2001 derived from the measurements of the MODIS on the NASA satellite *Terra*. Black indicates land and sea ice. The effects of cloud cover have been removed by compositing the many images of the cloud-free parts. This compositing inevitably blurs some of the sharp but dynamic boundaries such as the waves which occur along the boundaries of the cold areas extending westward along the equator in the Pacific and Atlantic oceans. The image was generated at NASA Goddard Space Flight center using atmospheric algorithms developed at the Rosenstiel School of the University of Miami.

The first instrument was launched in 1978 on the TIROS-N (Television Infrared Observation Satellite), and a continuous series of AVHRRs have flown on the US NOAA (National Oceanic and Atmospheric Administration) series of polar orbiting operational weather satellites since 1980. A third generation of the AVHRR was launched on the first of a new series of European Meteorological Operational (MetOp) satellite in 2006, with the second launched in 2012. The AVHRR is a relatively simple instrument with five (six in later models) spectral bands of which three ($\lambda = 3.7, 10.5$, and $11.5 \mu\text{m}$) are used for SST retrievals. The spatial resolution at the sea surface is $\sim 1 \text{ km}$ beneath the satellite, becoming larger toward the edges of the $\sim 2,700 \text{ km}$ wide swath. The swath is scanned by a continuously rotating plane mirror inclined at 45° to the axis of rotation that is aligned with the flight vector.

The first MODerate-resolution Imaging Spectroradiometer (MODIS (Esaias et al., 1998)) was launched on the NASA satellite *Terra* in 1999, with a second instrument being launched on *Aqua* in 2002. MODIS is a complex, 36-channel radiometer with five bands being used for SST retrievals. In addition to the standard two in the $10\text{--}12 \mu\text{m}$ window which use the NLSST atmospheric correction algorithm for SST derivation, there are three bands in the $3.7\text{--}4.1 \mu\text{m}$ window, which provide superior accuracy for nighttime SST retrievals. The design of MODIS includes many innovative features intended to improve the instrument performance over that achieved by AVHRR and the accuracy of the derived

variables including SST. These include a two-sided “paddle-wheel” scan mirror and 10 detectors per band (Figure 1).

The successor to the NOAA operational satellite series and the AVHRR will be the Joint Polar Satellite System (JPSS) and the Visible/Infrared Imager/Radiometer Suite (VIIRS). VIIRS is a 22-band radiometer with four bands to be used for SST with wavelengths of $3.7, 4.0, 10.8$, and $12.0 \mu\text{m}$, thereby offering the prospect of continuing the time series of SSTs into future decades. The first VIIRS was launched on the Suomi National Polar-orbiting Partnership (S-NPP) satellite on October 28, 2011.

An alternative approach of addressing the problem of correcting the effects of the atmosphere is to take two measurements of brightness temperatures through different atmospheric path lengths. This is the approach used by a series of European infrared radiometers, the Along-Track Scanning Radiometers (ATSR), first launched in 1991. The ATSRs use an inclined conical scan, rather than the linear cross-track scan used by the other instruments described above, which provides a measurement through one thickness of the atmosphere beneath the satellite and through about two thicknesses through the inclined view ahead of the satellite. The ATSR has the same infrared bands as AVHRR with wavelengths of $3.7, 10.8$, and $12.0 \mu\text{m}$, and the linear combination of the brightness temperatures in these bands at two angles provides a more accurate SST retrieval than without the dual view (Noyes et al., 2006). Again, the $3.7 \mu\text{m}$ measurements can only be used during the nighttime part of each orbit.

The disadvantage of the conical scan is the limited width of the swath, being about 500 km centered on the subsatellite track (Prata et al., 1990; Minnett, 1995). The third instrument of this type, the Advanced ATSR, flew on *Envisat* and provided data for a decade before mission ended on 8 April 2012, following an irreversible loss of communications with the satellite.

Dual-view SST observations are planned to resume with the launch of the Sentinel-3 missions, the first of which is planned for 2014 (Donlon et al., 2012). These will be the Sea and Land Surface Temperature Radiometers (SLSTRs (Coppo et al., 2010)), which are similar to the ATSRs but will have a wider swath. Further improved coverage will be gained from the two satellites being operational at the same time.

Microwave radiometers

The atmospheric contamination of microwave brightness temperatures is primarily from water vapor, oxygen, and liquid water. While the vertically integrated oxygen absorption is effectively constant, the water vapor and liquid water are highly variable. The microwave radiometers are designed to measure in atmospheric windows for SST, but also at wavelengths that are strongly absorbing in water vapor or liquid water. This allows simultaneous measurement of these variables, which can then be used to determine SST through cloud and water vapor. Microwave radiometers require large mirrors (antennas) and have a much lower spatial resolution than infrared SST measurements. Furthermore, the ill-defined edges of the fields of view of the instruments that result from the diffraction limitation of the antenna means that measurements close to coastlines are contaminated by emission from the land surface through the antenna side lobes. This limits microwave measurements of SST to about 50–100 km offshore. The sea surface emissivity in the microwave is very dependent on the emission angle, and to minimize this effect on the brightness temperatures, the antennas rotate about a vertical axis tracing out a curved swath on the sea surface at constant zenith angle. Emissivity of the ocean surface is highly dependent on wind speed and direction, which can be estimated by measuring vertical and horizontal polarizations. A combination of the brightness temperatures results in the retrieval of SST, wind speed, and other geophysical variables (Wentz and Meissner, 2000).

Although earlier instruments were capable of measuring SSTs, the current time series of measurements began in 1997 with the launch of the TMI (TRMM Microwave Imager) on the Tropical Rainfall Measuring Mission (TRMM) satellite (Kummerow et al., 1998). As its name suggests, TRMM is focused on tropical measurements, and the SST measurements are restricted to within about 40° of the equator.

The *Aqua* satellite that carries a MODIS also has the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E; Kawanishi et al., 2003),

which is the current prime source of microwave SST measurements. The polar orbit removes the latitude constraint of the TMI. AMSR-E has six frequency channels from 6.9 to 89 GHz, with the low-frequency channel providing sensitivity to SST variations. The 6.9 GHz channel has a footprint of 43×75 km, but the data are oversampled and SST is retrieved on a spatial grid of 25 km across a swath of 1,450 km width. The AMSR-E measurements were terminated on October 2011 when the torque required to turn the antenna exceeded safe limits, presumably caused by the consumption of lubricant.

The next-generation AMSR, AMSR-2, was launched on the Japanese GCOM-W1 (Global Change Observation Mission – Water) satellite on May 17, 2012. The AMSR-2 has similar characteristics to AMSR-E but with a larger, 2 m antenna.

Applications

The SST is one of the most important parameters in the Earth's climate system. It is related to the heat stored in the upper ocean and helps regulate the flow of heat to the overlying atmosphere, thus linking the two great fluid components of the climate. The heat in the upper ocean feeds the growth of severe storms (Shay et al., 2000) or, where the heat content has been diminished by the recent passage of a prior storm (e.g., Minnett, 2007), can lead to sudden weakening of a hurricane (Wentz et al., 2000). Accurate SST fields are required as input to Numerical Weather Prediction (NWP) models and the emerging activity of operational oceanography (Donlon et al., 2007). Time series of global SST fields have been suggested as sensitive indicators of the response of the climate to greenhouse gas forcing (Allen et al., 1994; Good et al., 2007). The patterns of SST reveal the structure of surface currents, the complexity of the oceanic mesoscale eddies and frontal meanders, and the characteristics of upwelling areas, and perturbations of the SST fields indicate climate anomalies, such as the El Niño–Southern Oscillation (ENSO) that modifies weather patterns over a large part of the globe for months to seasons at a time (Philander, 1989).

Accuracies of the sea surface temperatures

The usefulness of SSTs in any application is determined by the accuracy of the retrievals. Too great an uncertainty in the derived SSTs would render them unsuitable for many purposes – given the small size of the signals being sought, an accuracy of a few tenths of a degree is required. Not only is this difficult to achieve, but it is also difficult to demonstrate that it has been achieved. The usual way of determining the uncertainties in the satellite measurements is to compare them with other data taken at the same time and place, and the most bountiful source of independent temperatures are the surface drifting buoys. Over 10,000 sets of coincident measurements of wide-swath satellite imagers and drifting buoys in cloud-free conditions can be expected in most years. These provide

confident estimates of the differences between the satellite and the bulk SSTs (e.g., Kilpatrick et al., 2001), but because the buoys measure at 0.1–1.0 m depth, these statistics include contributions from the temperature differences caused by variations in the skin effect and in diurnal heating. A more accurate determination of the satellite SST uncertainties involves the use of ship board infrared radiometers to measure the skin SST (Barton et al., 2004), but the number of coincident measurements is much smaller, especially for the narrow swath ATSR series (Noyes et al., 2006). Nevertheless, the use of radiometric skin SST measurements reduces the values of the satellite SST uncertainties, by as much as a factor of 2 (Kearns et al., 2000). The best estimates of the SST accuracies from the AATSR is about 0.15–0.30 K (standard deviation) with a bias of < 0.2 K (Noyes et al., 2006; O’Carroll et al., 2008). For the AVHRR and MODIS, the standard deviations are 0.3–0.5 K, again with small mean uncertainties. AMSR-E SST uncertainties are 0.4–0.5 K (Wentz et al., 2003) and somewhat higher for TMI (Gentemann et al., 2004).

In regional comparisons, the mean and standard deviations of uncertainties can deviate from the global statistics (e.g., Kearns et al., 2000; Gentemann et al., 2008; Merchant et al., 2009). Anomalous atmospheric conditions can lead to larger errors in infrared SST retrievals, such as those that result from the presence of volcanic aerosols high in the atmosphere (Reynolds, 1993; Blackmore et al., 2012).

Summary

The accurate remote sensing of SST from polar-orbiting spacecraft is one of the major accomplishments of the satellite era. A time series of global SSTs now extends for over two decades and provides a valuable resource for studies of the climate system and many other oceanographic and meteorological phenomena. The provision of global SSTs with only a few hours delay is now an accepted part of numerical weather prediction, and research and operational efforts are currently directed at improving the accuracy of the SST fields and how they can contribute to better weather and ocean forecasting (Donlon et al., 2007). The continued measurement of SST from satellites is planned through the next decade and longer with the US JPSS, the Japanese GCOM, and the European MetOp and GMES (Global Monitoring for Environment and Security) Sentinel-3 programs.

Bibliography

- Allen, M. R., Mutlow, C. T., Blumberg, G. M. C., Christy, J. R., McNider, R. T., and Llewellyn-Jones, D. T., 1994. Global change detection. *Nature*, **370**, 24–25.
- Barton, I. J., Minnett, P. J., Donlon, C. J., Hook, S. J., Jessup, A. T., Maillet, K. A., and Nightingale, T. J., 2004. The Miami 2001 infrared radiometer calibration and inter-comparison: 2. Ship comparisons. *Journal of Atmospheric and Oceanic Technology*, **21**, 268–283.
- Blackmore, T., O’Carroll, A., Fennig, K., and Saunders, R., 2012. Correction of AVHRR Pathfinder SST data for volcanic aerosol effects using ATSR SSTs and TOMS aerosol optical depth. *Remote Sensing of Environment*, **116**, 107–117.
- Coppo, P., Ricciarelli, B., Brandani, F., Delderfield, J., Ferlet, M., Mutlow, C., Munro, G., Nightingale, T., Smith, D., Bianchi, S., Nicol, P., Kirschstein, S., Hennig, T., Engel, W., Frerick, J., and Nieke, J., 2010. SLSTR: a high accuracy dual scan temperature radiometer for sea and land surface monitoring from space. *Journal of Modern Optics*, **57**, 1815–1830.
- Cracknell, A. P., 1997. *The Advanced Very High Resolution Radiometer*. London: Taylor and Francis.
- Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., and Murray, J., 2002. Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, **15**, 353–369.
- Donlon, C., Robinson, I., Casey, K. S., Vazquez-Cuervo, J., Armstrong, E., Arino, O., Gentemann, C., May, D., LeBorgne, P., Piollé, J., Barton, I., Beggs, H., Poulter, D. J. S., Merchant, C. J., Bingham, A., Heinz, S., Harris, A., Wick, G., Emery, B., Minnett, P., Evans, R., Llewellyn-Jones, D., Mutlow, C., Reynolds, R. W., Kawamura, H., and Rayner, N., 2007. The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bulletin of the American Meteorological Society*, **88**, 1197–1213.
- Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M. H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., and Sciarra, R., 2012. The global monitoring for environment and security (GMES) sentinel-3 mission. *Remote Sensing of Environment*, **120**, 37–57.
- Esaias, W. E., Abbott, M. R., Barton, I., Brown, O. B., Campbell, J. W., Carder, K. L., Clark, D. K., Evans, R. H., Hoge, F. E., Gordon, H. R., Balch, W. M., Letelier, R., and Minnett, P. J., 1998. An overview of MODIS capabilities for ocean science observations. *IEEE Transactions on Geosciences and Remote Sensing*, **36**, 1250–1265.
- Gentemann, C. L., and Minnett, P. J., 2008. Radiometric measurements of ocean surface thermal variability. *Journal of Geophysical Research*, **113**, C08017. doi:10.1029/2007JC004540.
- Gentemann, C. L., Wentz, F. J., Mears, C. A., and Smith, D. K., 2004. In situ validation of Tropical Rainfall Measuring Mission microwave sea surface temperatures. *Journal of Geophysical Research*, **109**, C04021. doi:10.1029/2003JC002092.
- Gentemann, C. L., Minnett, P. J., LeBorgne, P., and Merchant, C. J., 2008. Multi-satellite measurements of large diurnal warming events. *Geophysical Research Letters*, **35**, L22602.
- Good, S. A., Corlett, G. K., Remedios, J. J., Noyes, E. J., and Llewellyn-Jones, D. T., 2007. The global trend in sea surface temperature from 20 years of advanced very high resolution radiometer data. *Journal of Climate*, **20**, 1255–1264.
- Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H., and Spencer, R. W., 2003. The advanced microwave scanning radiometer for the earth observing system (AMSR-E), NASDA’s contribution to the EOS for global energy and water cycle studies. *Geoscience and Remote Sensing, IEEE Transactions on*, **41**, 184–194.
- Kearns, E. J., Hanafin, J. A., Evans, R. H., Minnett, P. J., and Brown, O. B., 2000. An independent assessment of Pathfinder AVHRR sea surface temperature accuracy using the Marine-Atmosphere Emitted Radiance Interferometer (M-AERI). *Bulletin of the American Meteorological Society*, **81**, 1525–1536.
- Kilpatrick, K. A., Podestá, G. P., and Evans, R. H., 2001. Overview of the NOAA/NASA Pathfinder algorithm for sea surface temperature and associated matchup database. *Journal of Geophysical Research*, **106**, 9179–9198.

- Kummerow, C., Barnes, W., Kozu, T., Shiue, J., and Simpson, J., 1998. The tropical rainfall measuring mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, **15**, 809–817.
- Llewellyn-Jones, D. T., Minnett, P. J., Saunders, R. W., and Závody, A. M., 1984. Satellite multichannel infrared measurements of sea-surface temperature of the N.E. Atlantic ocean using AVHRR/2. *Quarterly Journal of the Royal Meteorological Society*, **110**, 613–631.
- May, D. A., Parmeter, M. M., Olszewski, D. S., and McKenzie, B. D., 1998. Operational processing of satellite sea surface temperature retrievals at the naval oceanographic office. *Bulletin of the American Meteorological Society*, **79**, 397–407.
- Merchant, C. J., and LeBorgne, P., 2004. Retrieval of sea surface temperature from space, based on modeling of infrared radiative transfer: capabilities and limitations. *Journal of Atmospheric and Oceanic Technology*, **21**, 1734–1746.
- Merchant, C. J., Harris, A. R., Maturi, E., and MacCallum, S., 2005. Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quarterly Journal of the Royal Meteorological Society*, **131**, 2735–2755.
- Merchant, C.J., Harris, A.R., Roquet, H., & Le Borgne, P., 2009. Retrieval characteristics of non-linear sea surface temperature from the Advanced Very High Resolution Radiometer. *Geophysical Research Letters*, **36**, L17604. doi:10.1029/2009GL039843.
- Minnett, P. J., 1995. The along-track scanning radiometer: instrument details. In Ikeda, M., and Dobson, F. (eds.), *Oceanographic Applications of Remote Sensing*. Boca Raton: CRC Press, pp. 461–472.
- Minnett, P. J., 2007. Heat in the ocean. In King, M. D., Parkinson, C. L., Partington, K. C., and Williams, R. G. (eds.), *Our Changing Planet, the View from Space*. Cambridge: Cambridge University Press, pp. 156–161.
- Minnett, P. J., Smith, M., and Ward, B., 2011. Measurements of the oceanic thermal skin effect. *Deep Sea Research Part II: Topical Studies in Oceanography*, **58**, 861–868.
- Noyes, E. J., Minnett, P. J., Remedios, J. J., Corlett, G. K., Good, S. A., and Llewellyn-Jones, D. T., 2006. The accuracy of the AATSR sea surface temperatures in the Caribbean. *Remote Sensing of Environment*, **101**, 38–51.
- O’Carroll, A. G., Eyre, J. R., and Saunders, R. W., 2008. Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *Journal of Atmospheric and Oceanic Technology*, **25**, 1197–1207.
- Philander, S. G. (ed.), 1989. *El Niño, La Niña, and the Southern Oscillation*. San Diego: Academic. International Geophysics Series.
- Prata, A. J. F., Cechet, R. P., Barton, I. J., and Llewellyn-Jones, D. T., 1990. The along-track scanning radiometer for ERS-1 – scan geometry and data simulation. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 3–13.
- Reynolds, R. W., 1993. Impact of Mount Pinatubo aerosols on satellite-derived sea surface temperature. *Journal of Climate*, **6**, 768–774.
- Saunders, R. W., and Kriebel, K. T., 1988. An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing*, **9**, 123–150. Correction, 1988, *International Journal of Remote Sensing*, **9**, 1393–1394.
- Shay, L. K., Goni, G. J., and Black, P. G., 2000. Effects of a warm oceanic feature on Hurricane Opal. *Monthly Weather Review*, **128**, 1366–1383.
- Walton, C. C., Pichel, W. G., Sapper, J. F., and May, D. A., 1998. The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *Journal of Geophysical Research*, **103**, 27999–28012.
- Wentz, F. J., and Meissner, T., 2000. *AMSR Ocean Algorithm (ATBD), Version 2*. Santa Rosa: Remote Sensing Systems.
- Wentz, F. J., Gentemann, C., Smith, D., and Chelton, D., 2000. Satellite measurements of sea-surface temperature through clouds. *Science*, **288**, 847–850.
- Wentz, F., Gentemann, C. L., and Ashcroft P., 2003. On-orbit calibration of AMSR-E and the retrieval of ocean products. In *12th Conference on Satellite Meteorology and Oceanography*, American Meteorological Society, Long Beach.

Cross-references

[Calibration, Microwave Radiometers](#)
[Calibration, Optical/Infrared Passive Sensors](#)
[Climate Data Records](#)
[Climate Monitoring and Prediction](#)
[Microwave Radiometers](#)
[Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties](#)
[Optical/Infrared, Radiative Transfer](#)
[Thermal Radiation Sensors \(Emitted\)](#)

SEA SURFACE WIND/STRESS VECTOR

W. Timothy Liu and Xiaosu Xie
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Definition

Wind is air in motion and is basically a vector quantity with a magnitude (speed) and a direction. Ocean surface stress is another vector quantity closely related to wind; it is the turbulent transfer of momentum between the ocean and the atmosphere. Their different applications are summarized in the section “[Scientific significance](#).”

Introduction

Sailors understand both the importance and the difficulty in obtaining information on wind over oceans. The textbooks still describe global ocean wind distribution in sailor’s terms: the calms of the Doldrums and Horse Latitudes, the steady Trade Winds, and the ferocity of the Roaring Forties. Just a few decades ago, almost all ocean wind measurements came from merchant ships. However, the quality and geographical distribution of these wind reports were uneven. Today, operational numerical weather prediction (NWP) also gives us wind information, but NWP depends on models, which are limited by our knowledge of the physical processes and the availability of data.

The ocean interacts with the atmosphere in nonlinear ways; processes at one scale affect processes at other scales. Adequate coverage can only be achieved from the vantage point of space. Space-based microwave sensors observe through clouds and measure ocean surface properties, backscatter, or emissivity, which are more

directly related to stress than wind. The microwave scatterometer is the best-established instrument to measure surface stress magnitude and direction under clear and cloudy conditions, night and day, as described in the section “[Scatterometry](#).” Until the launch of the scatterometer, our knowledge of stress distribution was largely derived from winds. Wind and stress have been mixed up in the interpretation of these space-based measurements. The section “[Wind and Stress Relation](#)” helps us to understand their relation and makes better use of the scatterometer observations.

The primary functions of the radar altimeter, the synthetic aperture radar (SAR), and the microwave radiometer are not wind-stress measurements, but they give wind speed as a secondary product. Wind speed, even without direction, is important, and wind speed from these sensors can be applied with directional information derived from other means. The polarimetric radiometer was also found recently to be sensitive to wind direction. The techniques of these sensors are briefly summarized in the section “[Other sensors](#).”

Scientific significance

Ocean wind is strongly needed for marine weather forecast and to avoid shipping hazards. Space-based wind measurements have been assimilated into operational NWP and used routinely in national centers for marine warning and forecasting. Surface wind convergence brings moisture and latent heat that drives deep convection and fuels hurricanes. The significance of wind measurement is clearly felt, for example, when a hurricane suddenly intensifies and changes course or when the unexpected delay of monsoon brings drought. Detailed distribution of wind power is also needed for the optimal deployment of floating wind farms in open sea that are enabled by new technologies (Liu et al., 2008a).

Wind-induced stress drives ocean current (ageostrophic component) and generates wave. The two-dimensional stress field is needed to compute the divergence and curl (vorticity) that control the vertical mixing. The mixing brings short-term momentum and heat trapped in the surface mixed layer into the deep ocean, where they are stored over time. It also brings nutrients and carbon stored in the deep ocean to the surface, where there is sufficient light for photosynthesis. The horizontal currents, driven in part by stress, distribute the stored heat and carbon in the ocean. Stress affects the turbulent transfer of heat, moisture, and gases between the ocean and the atmosphere and is critical in understanding and predicting weather and climate changes.

Relations between wind and stress

Ocean surface stress (τ) is the turbulent transfer of momentum generated by atmospheric instability caused both by wind shear (difference between wind and current) and buoyancy (vertical density stratification resulting from temperature and humidity gradients). Direct τ

measurement has only been done in a few field campaigns in the past. For all practical purposes, our knowledge of τ is derived from winds (U) at a reference height through a drag coefficient C_D , which is defined by

$$\tau = \rho C_D (U - U_s)^2 \quad (1a)$$

where U_s is the surface current and ρ is the air density.

In a similar fashion, the turbulent fluxes of heat (H) and moisture (E) have been related to the mean parameters – wind speed U , potential temperature T , specific humidity Q at 10 m, and [sea surface temperature](#) T_s , through

$$H = \rho C_P C_H (T - T_s)(U - U_s) \quad (2a)$$

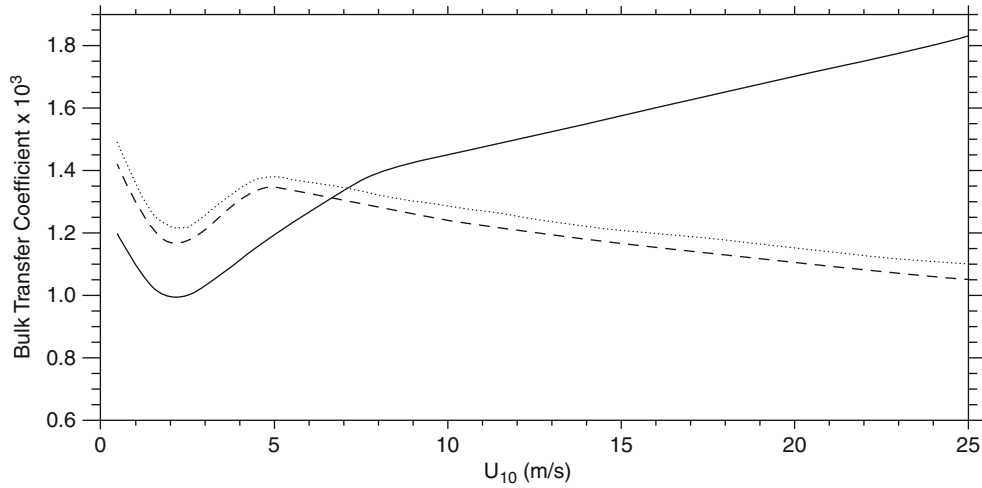
$$E = \rho C_E (Q - Q_s)(U - U_s) \quad (3a)$$

where C_P is the isobaric specific heat. The mean parameters are the measurements generally available from routine ship reports. In the past, the transfer coefficients, C_H and C_E , were approximated with the same values as C_D .

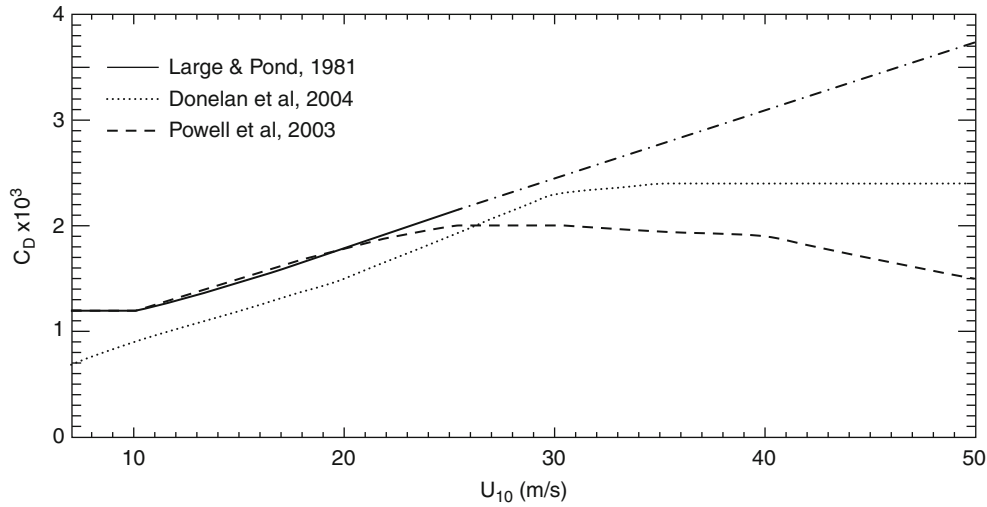
The transfer coefficient has been derived largely in field campaigns. [Figure 1](#), from Liu et al. (1979), illustrates the behavior of C_D at neutral stability. At low wind speed ($U < 3$ m/s), the flow is smooth; C_D increases with decreasing wind speed. And at moderate wind $3 < U < 25$ m/s, C_D is an increasing function of wind speed for a rough sea with open fetch. Stability and surface roughness are the main factors affecting the variability of the coefficients. Secondary factors, such as sea states and spray from breaking waves, whose data are not generally available are not included in these parameterization schemes and should be part of the errors (e.g., Bourassa et al., 1999).

Liu et al. (1979) first postulated that, in a rough sea, under a moderate range of winds, C_H and C_E do not increase with wind speed because of molecular constraint at the interface, while C_D may still increase because momentum is transported by form drag. Liu’s hypothesis, as illustrated in [Figure 1](#), was subsequently supported by measurements in field experiments (e.g., DeCosmo et al., 1996).

Emanuel (1995) argued, from theoretical and numerical model results, that the scenario of Liu et al. (1979) could not hold at the strong wind regime of a hurricane. To attain the wind strength of a hurricane, the energy dissipated by drag could not keep increasing while the energy fed by sensible and latent heat does not increase with wind speed. His argument puts limit on the increase of C_D as a function of wind speed. The postulation of the level-off of the increase of C_D with wind speed at hurricane scale winds was supported by the results of the laboratory studies of Donelan et al. (2004) and the aircraft experiments by Powell et al. (2003), as illustrated in [Figure 2](#). The result of Large and Pond (1981), derived for the range of moderate wind speeds, is extrapolated to the range of strong wind speeds for comparison in the figure. The high wind behavior of C_D may post constrain in retrieving hurricane scale winds by the scatterometer.



Sea Surface Wind/Stress Vector, Figure 1 Variation of the bulk transfer coefficients of momentum (drag coefficient), heat, and moisture with wind speed by Liu et al. (1979).



Sea Surface Wind/Stress Vector, Figure 2 Variation of the drag coefficients in strong winds.

Liu et al. (1979), for the first time, performed the bulk parameterization of the surface fluxes by solving the nondimensional flux-profile relations (also called similarity functions), thus including the effects of stability and surface molecular constraints. The functions are:

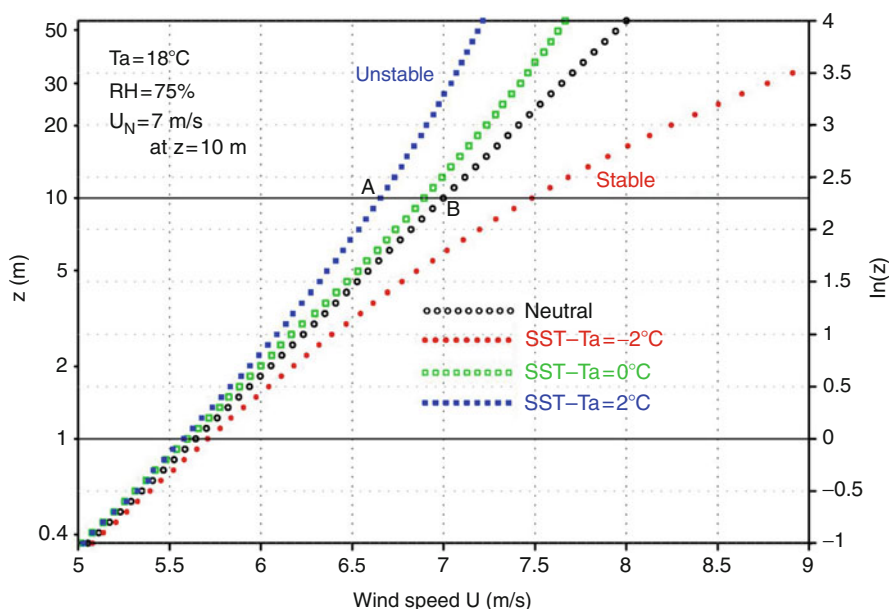
$$\frac{U - U_s}{U_*} = 2.5 \left(\ln \frac{z}{z_0} - \psi_U \right) = \frac{1}{\sqrt{C_D}} \quad (1b)$$

$$\frac{T - T_s}{T_*} = 2.5 \left(\ln \frac{z}{z_T} - \psi_T \right) = \frac{\sqrt{C_D}}{C_H} \quad (2b)$$

$$\frac{Q - Q_s}{Q_*} = 2.5 \left(\ln \frac{z}{z_Q} - \psi_Q \right) = \frac{\sqrt{C_D}}{C_E} \quad (3b)$$

where $U_* = (\tau/\rho)^{1/2}$ is the frictional velocity; T_* and Q_* are nondimensional flux parameters defined as $T_* = H/(C_P U_*)$ and $Q_* = E/(U_*)$; z_0 , z_T , and z_Q are the roughness lengths; and ϕ_U , ϕ_T , and ϕ_Q are functions of the stability parameter, which is the ratio of buoyancy to shear production of turbulence. When U_s and ϕ_U are zero, U becomes the equivalent neutral wind (U_N), which is uniquely related to U_* (or τ). U_N has been used as the geophysical product of the scatterometer.

Typical wind profiles at various stabilities are shown in Figure 3. The atmosphere is rarely exactly neutral. Even when there is no vertical temperature gradient, there is likely to be a humidity gradient. To compute U_N from conventional wind measurements of U (A on the blue



Sea Surface Wind/Stress Vector, Figure 3 Typical wind profiles at various stability conditions derived from the flux-profile relation by Liu et al. (1979). B is the equivalent neutral wind corresponding to the actual wind measurement at A.

curve in Figure 3), U_* and z_0 are computed as the slope and intercept at the surface of the curve. The neutral relation (straight line) defined by U_* and z_0 will give U_N (point B). This method has been used in algorithm development and calibration of all scatterometers launched by NASA. At a given level, U_N is larger than the actual wind (U) under unstable conditions but lower under stable condition. From Equation 1b, $U_N - U = 2.5U_*\psi_U$, assuming the dependence of z_0 on buoyancy is much smaller than on wind shear. This difference is the inherent error of using scatterometer measurement as the actual wind. While the parameterization method by Liu et al. (1979) has been improved (e.g., Fairall et al., 1996), the formulation of ψ was largely based on experimental data taken on land, validated only with small amount of measurements over ocean, and may still have considerable uncertainties.

Scatterometry

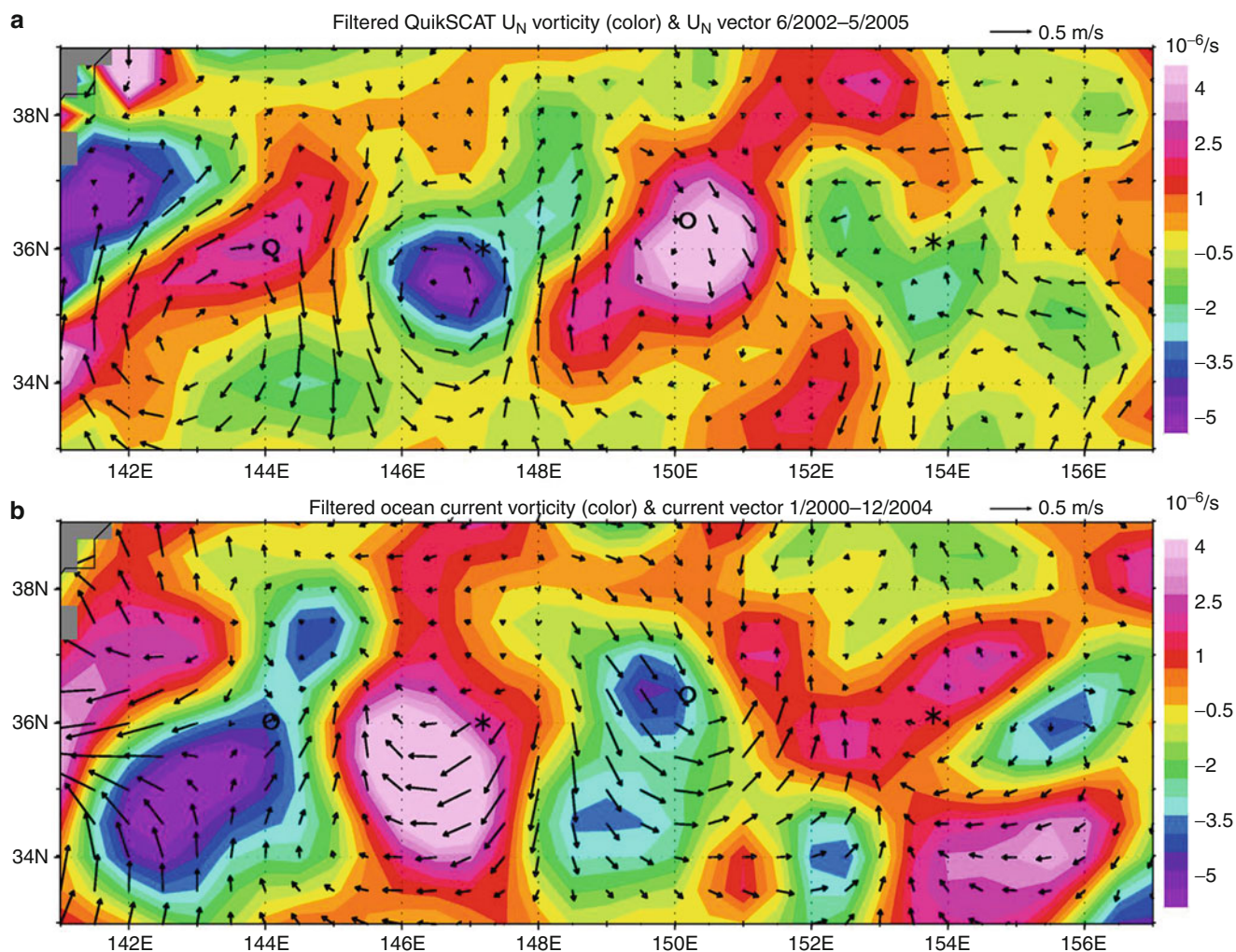
During the Second World War, marine radar operators observed noises on their radar screens, which obscured small boats and low-flying aircraft. They termed this noise “sea clutter.” This clutter was the backscatter of the radar pulses by the small waves on the ocean’s surface. The radar operators at that time were quite annoyed by this noise, not knowing that a few decades later, scientists would make important use of it.

Scatterometers send microwave pulses to the Earth’s surface and measure the power backscattered from the surface roughness. The roughness may describe the characteristics of polar ice or vegetation over land. Over the ocean, which covers over three quarters of the Earth’s

surface, the surface roughness is largely due to the small centimeter wavelength waves on the surface. These surface waves are believed to be in equilibrium with the local stress. The backscatter depends not only on the magnitude of the stress but also the stress direction relative to the direction of the radar beam (azimuth angle). The capability of measuring both stress magnitude and direction is the major, important characteristic of the scatterometer. Liu and Large (1981) were first to relate scatterometer and direct stress measurements.

There are far fewer stress than wind measurements for validation and calibration, and the equivalent neutral wind (U_N) is used as the geophysical product of the scatterometer (Liu and Tang, 1996). By definition, U_N is uniquely related to τ , while the relation between τ and U depends on atmospheric stability, as discussed in the previous section. U_N is similar to U_* and can be viewed as stress in wind units.

At incident angles greater than 20° , the radar return is governed by Bragg scattering, and the backscatter increases with U_N . The backscatter is governed by the inphase reflections from surface waves. The geophysical model function (GMF), from which U_N is retrieved from the observed backscatter, in form of the normalized radar cross section (σ_o), is largely based on empirical fits of data. The symmetry of backscatter with respect to wind direction requires observations at multiple azimuth angles to resolve the directional ambiguity. Because of the uncertainties in the wind retrieval algorithm and noise in the backscatter measurements, the problem with directional ambiguity was not entirely eliminated even with three



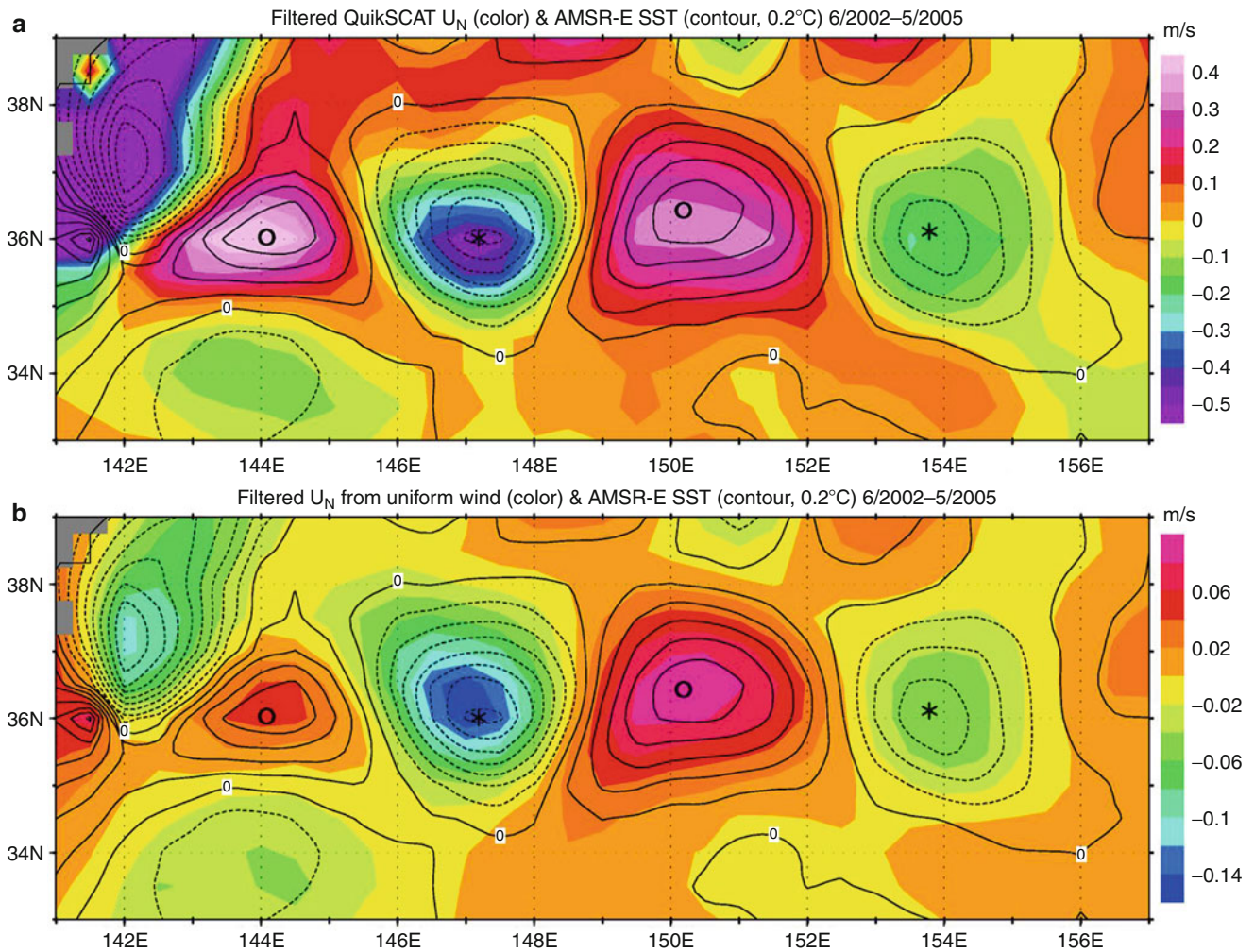
Sea Surface Wind/Stress Vector, Figure 4 (a) Filtered vector (*black arrows*) superimposed on vorticity (color, 10^{-6} s^{-1}) of U_N observed by QuikSCAT. (b) Filtered vector (*black arrows*) superimposed on vorticity of the surface current measured by Lagrangian drifters.

azimuthal looks in the scatterometers launched after Seasat. A median filter iteration technique has been commonly used to remove the directional ambiguity.

Although scatterometers are known to measure surface stress, they have been used and promoted as wind measuring instruments and U_N has been used as the actual wind, particularly in operational weather applications. This is justified on the assumption that, over the large expanse of ocean, ocean current is weak compared with wind and the atmosphere is under near neutral conditions. Over the Agulhas Extension and the Kuroshio Extension current with sharp horizontal current shear and temperature gradients, however, Liu et al. (2007) and Liu and Xie (2008) show that stress could be very different from winds. Over the current meanders, one would expect that the wind will be dragged in the direction of the current. Figure 4 clearly shows that where the vorticity of U_N measured by

QuikSCAT is positive, the vorticity of the surface current measured by the drifters is negative and vice versa, indicating almost opposite rotations. This is a clear indication that the scatterometer measurement has the characteristic of turbulent stress generated by shear rather than wind. The directional difference between scatterometer measurement and current exists because the scatterometer measures stress, which is the vector difference between wind and current (Park et al., 2006).

The ubiquitous spatial coherence between sea surface temperature (T_s) and U_N measured by the scatterometer found under a variety of atmospheric conditions is also the characteristics of turbulent stress generated by buoyancy. In the unstable region, atmospheric buoyancy generates turbulent momentum transport and increases the stress magnitude. Figure 5a shows the coherence over the Kuroshio Extension. Figure 5b shows similar



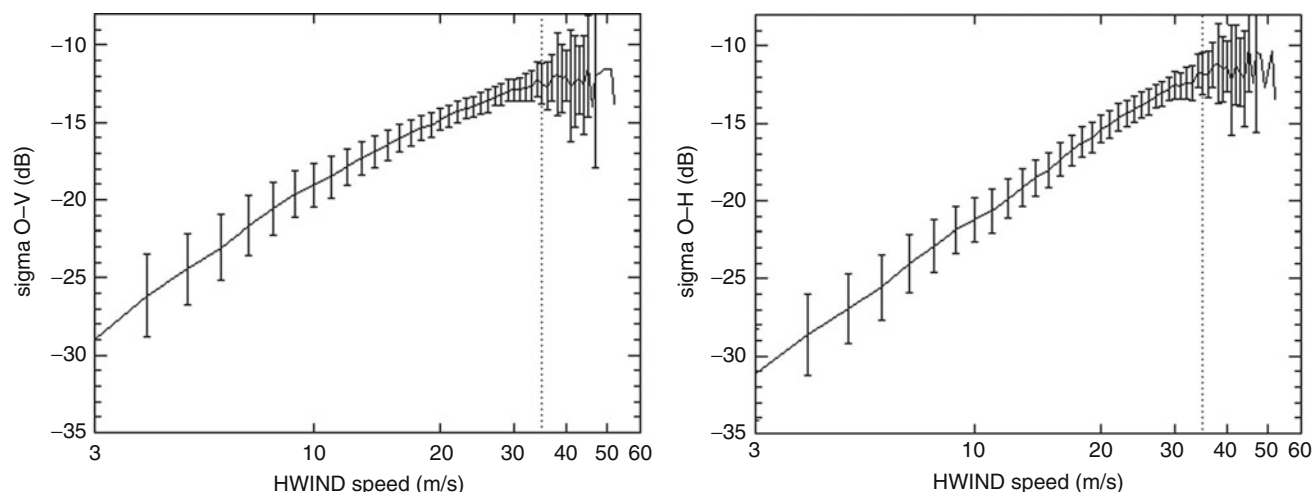
Sea Surface Wind/Stress Vector, Figure 5 Isotherms of filtered T_s measured by AMSR-E (0.2 °C interval) superimposed on (a) filtered magnitude of QuikSCAT U_N (color, m/s), and (b) filtered U_N computed from a uniform wind field of $u = 7.5$ m/s. Solid and broken lines represent positive and negative values, respectively.

coherence between T_s and U_N computed from a uniform wind field under similar stability condition, demonstrating that the coherence is the characteristic of stress and not wind. Factors affecting larger-scale wind, such as pressure gradient force, Coriolis force, and baroclinicity, are not important at the small scales of turbulence, and that is the reason of ubiquitous coherence under various atmospheric conditions. The higher stress over warmer water affects atmospheric wind aloft, but the influence will be subjected to these large-scale factors. Ocean parameters, such as surface current and temperature, are needed to derive wind from stress in these frontal regions.

Retrieving strong winds from the scatterometer measurements is also difficult. The problem is obvious in Figure 6, derived from NASA's scatterometer on QuikSCAT measuring at Ku-band. Data for the 12 hurricanes in the North Atlantic in the 2005 seasons, excluding those with over 10 % chances of rain, were examined.

Figure 6 shows that, in moderate winds ($U < 35$ m/s), the logarithm of σ_o (in db) increases linearly with the logarithm of wind speed, at both polarizations. At strong winds ($U > 35$ m/s), however, σ_o increases at a much slower rate with increasing wind speed. Similar saturation is found in the European Advanced Scatterometer (ASCAT), measuring at C-band. Such high wind saturation has also been observed from aircraft flying over hurricanes.

When the model function developed over the moderate wind range is applied to the strong winds, an underestimation of wind speed results. Strong efforts have been made to adjust the model function (slope in Figure 3) in strong winds and to find the right channel (combination of polarization, frequency, incident angle) that would be sensitive to the increase of strong winds (e.g., Fernandez et al., 2006). The success would be difficult if flow separation occurs at high winds and the surface roughness and stress



Sea Surface Wind/Stress Vector, Figure 6 Normalized radar cross section at two polarizations measured by QuikSCAT for 12 hurricanes as a function of colocated surface wind provided by the National Hurricane Center.

do not increase with winds as discussed in section “[Relations between wind and stress.](#)”

Direct stress retrieval

Weissman and Graber (1999) made an early attempt to build a GMF to retrieve stress instead of wind. There are many reasons for a GMF-S (stress) to retrieve stress (or U_*) directly rather than the present GMF-W to retrieve U_N . The first reason is the deficiency of the present GMF-W, which should be developed and calibrated with U_N computed from research quality in situ wind measurements, as described in the section “Relations between wind and stress.” Such computation of U_N from in situ measurements was performed before credible ocean surface wind products became available from operational NWP centers. Most of the tuning of the revised GMF after Seasat was based on NWP products that are not U_N (not corrected for stability dependence). The resultant errors are not reversible and difficult to gauge.

The second reason is the uncertainty of the drag coefficient. Ideally, stress could be derived from U_N retrieved from the scatterometer, using a neutral drag coefficient. However, if the drag coefficient is not the same as that was used to derive U_N for development of the GMF, error will be introduced through the uncertainty of drag coefficient.

The other two reasons are related to the directional difference between wind and stress. The procedure to “select” the stress direction should be different from wind direction in two ways. In the first way, we should initialize the ambiguity removal process with “nudging” fields that are more relevant to stress than wind. Where a strong ocean current exists, the stress should point to the direction of the vector difference between wind and current. The second way is to develop a flexible median filter to accommodate the small spatial scale of stress as compared

with winds. One of the problems of the selection is the lack of sufficient current information.

One of the reasons usually given for promoting scatterometer as a wind sensor instead of a stress sensor is that there is more wind than stress measurements to develop and calibrate the GMF. Such an explanation is not valid because U_N , by definition, has an unambiguous relation with stress and needs stress for computation. To provide each U_N for development or calibration of the GMF from measured wind U , stress or U_* has to be computed first as discussed in the section “Relations between wind and stress.”

The stress derived from wind in such a way is not ideal because it addresses only the stability problem but does not include current information. Such deficiency may be somewhat alleviated through the ambiguity removal process by using more appropriate filter size and nudging with the vector difference between wind and optimal surface current information that is available. The direct retrieval of stress depends not only on the fast and large-scale atmospheric circulation but also on the small-scale and slow ocean processes, as reflected in surface current and temperature.

Other sensors

Both the microwave altimeter and SAR are similar to the scatterometer in the sense that they are active sensors that send microwave pulses to the Earth’s surface and measure the backscattered power. The microwave radiometer is a passive sensor, observing the radiance emitted by the Earth and its atmosphere.

While the scatterometer views at oblique angles, the altimeter views at nadir (very small incident angles). At nadir, the backscattered energy is a result of specular reflection (the wavelets serve as small mirrors), and the backscatter is not sensitive to U_N direction. Because the

instrument is not scanning, data are only available at very narrow (2 km) repeated ground tracks. The coverage of all the past altimeters is poor compared with the scatterometer and the microwave radiometers.

A SAR looks perpendicular to aircraft path only at one azimuth angle and cannot resolve the U_N direction like the scatterometer. SAR has spatial resolutions that are much better than scatterometers, but the high resolution also introduces higher uncertainties in accuracy caused by secondary effects that affect surface roughness. The instrument and the data processing procedure are much more complicated than the scatterometer. The scatterometer GMF can be used to relate the σ_o measured by SAR to U_N . However, a particular value of σ_o may correspond to a range of U_N , depending on the azimuth angle. Hence, in order to retrieve U_N with the GMF, the U_N direction must first be specified. Whether the a priori direction information is derived from the orientation of km-scale structure in the SAR image or from operational NWP models, the spatial scales are much coarser than those of σ_o .

Ocean surface wind speed has also been derived from the radiance measured by a microwave radiometer. It is generally believed that wind speed affects the surface emissivity indirectly through the generation of ocean waves and foam. Radiometers designed to observe the ocean surface operate primarily at window frequencies, where atmospheric absorption is low. Radiances at frequencies sensitive to [sea surface temperature](#), atmospheric water vapor, and liquid water are also measured; they are used to correct for the slight interference by the atmosphere. It was demonstrated in several airborne experiments (e.g., Yueh et al., 1997) that the polarization properties of the sea surface emission vary not only as a function of the wind speed but also as a function of wind direction. The wind direction measuring capability is being tested by a polarimetric radiometer, WindSat on the Coriolis satellite, developed by the US Navy.

Potential improvements

Historically, the European Space Agency used the C-band (5 GHz), but NASA prefers the Ku-band (14 GHz) for their scatterometers. The backscatter at higher frequencies is more sensitive to shorter ocean waves. The Ku-band is more sensitive to weak wind-stress variations but is more subject to atmospheric effects and rain contamination. Attempts have been made to retrieve winds from L-band (1.4 GHz) scatterometer on Aquarius (Yueh et al., 2013). There have been calls for a multifrequency scatterometer that is sensitive to various parts of ocean surface wave spectrum and may reduce atmospheric and rain effects.

Present scatterometers are real-aperture systems and the spatial resolution is limited by the antenna size. A larger antenna will, of course, enhance the spatial resolution. Another way to achieve higher resolution is to add

a synthetic aperture capability. One of the drawbacks of present scatterometers is the ambiguity in retrieving wind-stress direction. The backscatter is a cosine function of the azimuth angle. The correlation between co-polarized and cross-polarized backscatter is a sine function of azimuth angle. By adding polarized measurement capability to the scatterometer, the directional ambiguity problem could be mitigated.

One polar-orbiting scatterometer at a low-altitude (e.g., 800 km) orbit can sample at a location on Earth not more than two times a day. Additional instrument flying in tandem will allow the description of higher temporal variability and the reduction of the aliasing (bias introduced by subsampling) of the mean wind stress. As demonstrated by Liu et al. (2008b), adding ASCAT to QuikSCAT decreases the time required to cover 90 % of the Earth from 24 to 19 h. Adding Oceansat-2 of India resolves the inertial frequency desired by oceanographers and provides 6 hourly repeated coverage required by operational weather forecasters. The adverse effect due to the demise of QuikSCAT should be mitigated by data from WindSat or the scatterometer on Haiyang-2 of China. The scatterometer on Space Station will provide new sampling opportunity.

Not all space-based ocean surface wind and stress measurements are comparable in quality. Standardizing the technology requirements for observation accuracies of different research and operational applications and international cooperation are very desirable. Many scientific reports have affirmed the need for high-quality, continuous, and consistent long time series of ocean surface vector winds and stress.

Conclusion

The basics of scatterometry and air-sea turbulence transfer are reviewed to bring out the uniqueness of the scatterometer in measuring ocean surface stress in addition to wind. The ubiquitous spatial coherence of scatterometer measurements with ocean surface temperature and current is attributed to the two ocean factors that drive the buoyancy and wind-shear production of turbulence transfer (stress); the factors are less directly influential on wind. The reduced sensitivity of the scatterometer to the increase of wind speed at hurricane scale winds is related to the failure of conventional drag coefficient caused by flow separation. A scatterometer that measures stress better than wind is still important to the estimation of the dynamic forcing and the oceanic feedback that affects the maintenance and intensification of the storm. The feasibility, advantage, and need for a geophysical model function to retrieve stress directly rather than the equivalent neutral wind (the present geophysical product of the scatterometer) are explained. The direct retrieval of stress from scatterometer measurements will enable new science applications with a new perspective, as expounded by Liu et al. (2010).

Acknowledgments

This study was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). It was jointly supported by the Ocean Vector Winds, Ocean Surface Salinity, and Precipitation Measuring Mission Programs of NASA. Wenqing Tang provided valuable assistance.

Bibliography

- Bourassa, M. A., Vincent, D. G., and Wood, W. L., 1999. A flux parameterization including the effects of capillary waves and sea state. *Journal of the Atmospheric Sciences*, **56**, 1123–1139.
- DeCosmo, J., Katsaros, K. B., Smith, S. D., Anderson, R. J., Oost, W. A., Bumke, K., and Chadwick, H., 1996. Air-sea exchange of water vapor and sensible heat: the humidity exchange over the sea (HEXOS) results. *Journal of Geophysical Research*, **101**, 12001–12016.
- Donelan, M. A., Haus, B. K., Reul, N., Plant, W. J., Stiassnie, M., Graber, H. C., Brown, O. B., and Saltzman, E. S., 2004. On the limiting aerodynamic roughness of the ocean in very strong winds. *Geophysical Research Letters*, **31**, L18306, doi:10.1029/2004GL019460.
- Emanuel, K., 1995. Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady state model incorporating eye dynamics. *Journal of the Atmospheric Sciences*, **52**, 3969–3976.
- Fairall, C. W., Bradley, E. F., Rogers, D. P., Edson, J. B., and Young, G. S., 1996. Bulk parameterization of air-sea fluxes in TOGA COARE. *Journal of Geophysical Research*, **101**, 3747–3767.
- Fernandez, D. E., Carswell, J. R., Frasier, S., Chang, P. S., Black, P. G., and Marks, F. D., 2006. Dual-polarized C- and Ku-band ocean backscatter response to hurricane-force winds. *Journal of Geophysical Research*, **111**, C08013, doi:10.1029/2005JC003048.
- Large, W. G., and Pond, S., 1981. Open ocean momentum flux measurements in moderate to strong winds. *Journal of Physical Oceanography*, **11**, 324–336.
- Liu, W. T., and Large, W. G., 1981. Determination of surface stress by Seasat-SASS: a case study with JASIN data. *Journal of Physical Oceanography*, **11**, 1603–1611.
- Liu, W. T., and Tang, W., 1996. *Equivalent Neutral Wind*. Pasadena: JPL Publication 96–17, Jet Propulsion Laboratory. 16 pp.
- Liu, W. T., and Xie, X., 2006. Measuring ocean surface wind from space. In Gower, J. (ed.), *Remote Sensing of the Marine Environment*, 3rd edn. Bethesda: American Society for Photogrammetry and Remote Sensing. Manual of Remote Sensing, Vol. 6, pp. 149–178. Chapter 5.
- Liu, W. T., and Xie, X., 2008. Ocean–atmosphere momentum coupling in the Kuroshio extension observed from space. *Journal of Oceanography*, **64**, 631–637.
- Liu, W. T., Katsaros, K. B., and Businger, J. A., 1979. Bulk parameterization of air-sea exchanges in heat and water vapor including the molecular constraints at the interface. *Journal of the Atmospheric Sciences*, **36**, 1722–1735.
- Liu, W. T., Xie, X., and Niiler, P. P., 2007. Ocean–atmosphere interaction over Agulhas extension meanders. *Journal of Climate*, **20**, 5784–5797.
- Liu, W. T., Tang, W., and Xie, X., 2008a. Wind power distribution over the ocean. *Geophysical Research Letters*, **35**, L13808, doi:10.1029/2008GL034172.
- Liu, W. T., Tang, W., Xie, X., Navalgund, R., and Xu, K., 2008b. Power density of ocean surface wind-stress from international

scatterometer tandem missions. *International Journal of Remote Sensing*, **29**, 6085–6090.

- Liu, W. T., Xie, X., and Tang, W., 2010. Scatterometer’s unique capability in measuring ocean surface stress. In Barale, V., Gower, J. F. A., and Alberotanza, L. (eds.), *Oceanography from Space*. New York: Springer, pp. 93–111. Chapter 6.
- Park, K.-A., Cornillon, P., and Codiga, D. L., 2006. Modification of surface winds near ocean fronts: effects of the Gulf Stream rings on scatterometer (QuikSCAT, NSCAT) wind observations. *Journal of Geophysical Research*, **111**, C03021, doi:10.1029/2005JC003016.
- Powell, M. D., Vickery, P. J., and Reinhold, T., 2003. Reduced drag coefficient for high wind speeds in tropical cyclones. *Nature*, **422**, 279–283.
- Weissman, D. E., and Graber, H. C., 1999. Satellite scatterometer studies of ocean surface stress and drag coefficients using a direct model. *Journal of Geophysical Research*, **104**(C5), 11329–11335.
- Yueh, S. H., Wilson, W. J., Li, F. K., Nghiem, S. V., and Ricketts, W. B., 1997. Polarimetric brightness temperatures of sea surfaces measured with aircraft K- and Ka-band radiometers. *IEEE Transactions on Geoscience and Remote Sensing*, **35**, 1177–1185.
- Yueh, S. H., Tang, W., Fore, A. G., Neumann, G., Hayashi, A., Freedman, A., Chaubell, J., & Lagerloef, G. S. E., 2013. L-band passive and active microwave geophysical model functions of ocean surface winds and applications to Aquarius retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, **51**, 4619–4632. See http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6553597&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D6553597.

Cross-references

Climate Data Records
Microwave Surface Scattering and Emission
Radar, Scatterometers
Radiation, Polarization, and Coherence
Severe Storms
Tropospheric Winds
Water and Energy Cycles

SEVERE STORMS

Charles A. III Doswell
Doswell Scientific Consulting, Norman, OK, USA

Synonyms

Severe convection; Severe local storms; Severe weather

Definition

Severe storms. Any deep convective storm, usually associated with lightning and thunder, that produces one or more of the following: large hailstones, strong winds, and/or tornadoes. For the United States, the threshold diameter for severe hail is 2.54 cm (1 in.) and the threshold speed for severe winds is 25 m s^{−1} (50 knots), but *any* tornado is considered severe (tornadoes are ranked in six categories according to the *Enhanced Fujita scale* (Fujita, 1971), ranging from EF0 to EF5). In many countries,

heavy rainfall is also considered a severe storm, with varying thresholds for rainfall per unit time (Doswell III et al. 2009).

Introduction: deep, moist convective storms

Any type of weather can be called severe: extremely high or low temperatures, extended periods without rainfall, heavy snowfall, freezing drizzle or rain, and so on. However, the generally accepted use of the term “severe weather” is for the events produced by deep convective storms (Ludlam, 1963; Doswell III, 2001). Such storms are typically accompanied by lightning and thunder caused by that lightning, but in some cases, the clouds are not sufficiently electrified to produce lightning discharges while still capable of some forms of severe weather.

Severe convection is the result of a condition broadly referred to as *conditional instability* (Schultz et al., 2000). This type of instability results from an excess of latent and sensible heat at low levels, which in turn is due to incoming solar radiation that both heats the surface and enhances evapotranspiration of water into the atmosphere in vapor form. The heated surface conducts its heat to the atmosphere immediately above the surface, which then carries that heat upward by *convection*. As meteorologists use the term, convection refers to the transport of atmospheric properties in the vertical by the motion of the air itself – it can be considered a form of mixing that is a response to the imbalance created when the air is heated from below.

Water vapor plays a critical role in the development of severe convection by releasing latent heat – which became “latent” when the liquid water evaporated into the overlying air or was “exhaled” during transpiration by plants – during the process of condensation into cloud droplets. It is the energy associated with latent heat release that powers deep convective clouds and which then can be manifest in the various forms of severe weather. Condensation of this water vapor into cloud droplets and, ultimately, into various forms of precipitation is the result of air containing sufficient amounts of water vapor being forced upward by some external process, such as a so-called front or by flowing up the side of a mountain. Pressure in the atmosphere decreases upward owing to the effect of gravity, so ascent results in expansion of the ascending air currents (*updrafts*). Surrounding air is drawn into the developing updraft, producing what is known as *inflow*. Following the laws of thermodynamics, the air’s expansion during its ascent results in cooling. For all practical purposes, it can be assumed that rising air exchanges no heat with its surroundings during its ascent, which means the process is *adiabatic*. The rate of cooling with height in the rising air prior to condensation is called the *dry adiabatic lapse rate*, corresponding to a constant value of $9.8\text{ }^{\circ}\text{C km}^{-1}$. During dry adiabatic ascent, the water vapor content of the rising air current remains constant, so that ultimately, the rising air reaches a level at which

the relative humidity becomes high enough for condensation to begin (near 100 % relative humidity, where the air is said to be *saturated* with water vapor). Condensation can begin slightly before attaining a relative humidity of 100 % owing to the presence of tiny particles called *condensation nuclei*, which promote condensation because of their hygroscopic properties.

Once condensation begins, the released latent heat causes the ascending air to cool at a lesser rate, known as the *moist adiabatic lapse rate*, but this value is not a constant. Rather, the rate is smallest at low levels (around $6\text{ }^{\circ}\text{C km}^{-1}$) and approaches the dry adiabatic rate asymptotically as the air continues to rise. Because of the release of latent heat, the ascending air may become less dense than its surroundings (when temperature decreases with height sufficiently in that environment) and so becomes positively *buoyant*. If the lapse rate in the environment is between the dry and moist adiabatic values, then that environment is considered conditionally unstable – it is stable for ascent without condensation, but can become unstable when the ascending air becomes saturated. Positive buoyancy results in a strong acceleration of the air’s ascent – it has become unstable and continues to rise on its own after having been lifted initially by some external mechanism. In extreme cases of this updraft instability, the vertical speed can reach values as high as 50 m s^{-1} or more during an ascent to 10 km or beyond.

As the air in the updraft continues to rise, it eventually reaches a level where it loses its buoyancy, primarily because most of the moisture has condensed and its lapse rate increases to approach the dry adiabatic value. The height of this level also depends on the vertical temperature profile in the environment. Owing to the inertia of its ascent accumulated during its buoyant acceleration, it will overshoot this *equilibrium level* before dropping back downward and spreading out in the upper parts of the storm, forming the so-called anvil cloud (Figure 1). The cloud material the storm produces eventually will accumulate at the equilibrium level and ultimately re-evaporate.

Various processes within the developing convective cloud cause the conversion of cloud droplets, which are initially very small (perhaps only $20\text{ }\mu\text{m}$ in diameter) into precipitation-size particles (a few mm in diameter), which are large enough to have terminal velocities sufficient to let them fall out of the cloud (a few m s^{-1}). The development of precipitation within a convective cloud (Lamb, 2001) signals a transition in the evolution of a convective storm. The presence of precipitation results in the creation of descending air currents, or *downdrafts* through two physical processes. When precipitation falls into unsaturated environmental air, it begins to evaporate. Evaporation causes that environmental air to cool (by absorbing the latent heat of vaporization), so it can become negatively buoyant and begin to sink. Further, the physical presence of precipitation particles exerts a drag on the surrounding air, causing that air to descend



Severe Storms, Figure 1 A deep convective storm cloud (called a *cumulonimbus* cloud), with its anvil spreading out at the equilibrium level and being carried by the upper-level winds from *left to right*. The *top* of an overshooting updraft is visible above the anvil (compare to [Figure 2b](#)) (Photograph © 1999 C. Doswell).

along with the precipitation. Downdrafts of precipitation-cooled air eventually reach the surface and spread out, forming pools of cold, stable air. The leading edge of the *outflow* caused by downdrafts is called a *gust front*, as it is usually associated with gusty surface winds.

Only part of the water vapor condensed into cloud droplets ultimately falls out as precipitation. The fraction of the input water vapor deposited on the ground is known as the *precipitation efficiency*. Some storms are very inefficient, with only 10 % or less of the input water vapor falling out as precipitation. Others have efficiencies exceeding 50 %.

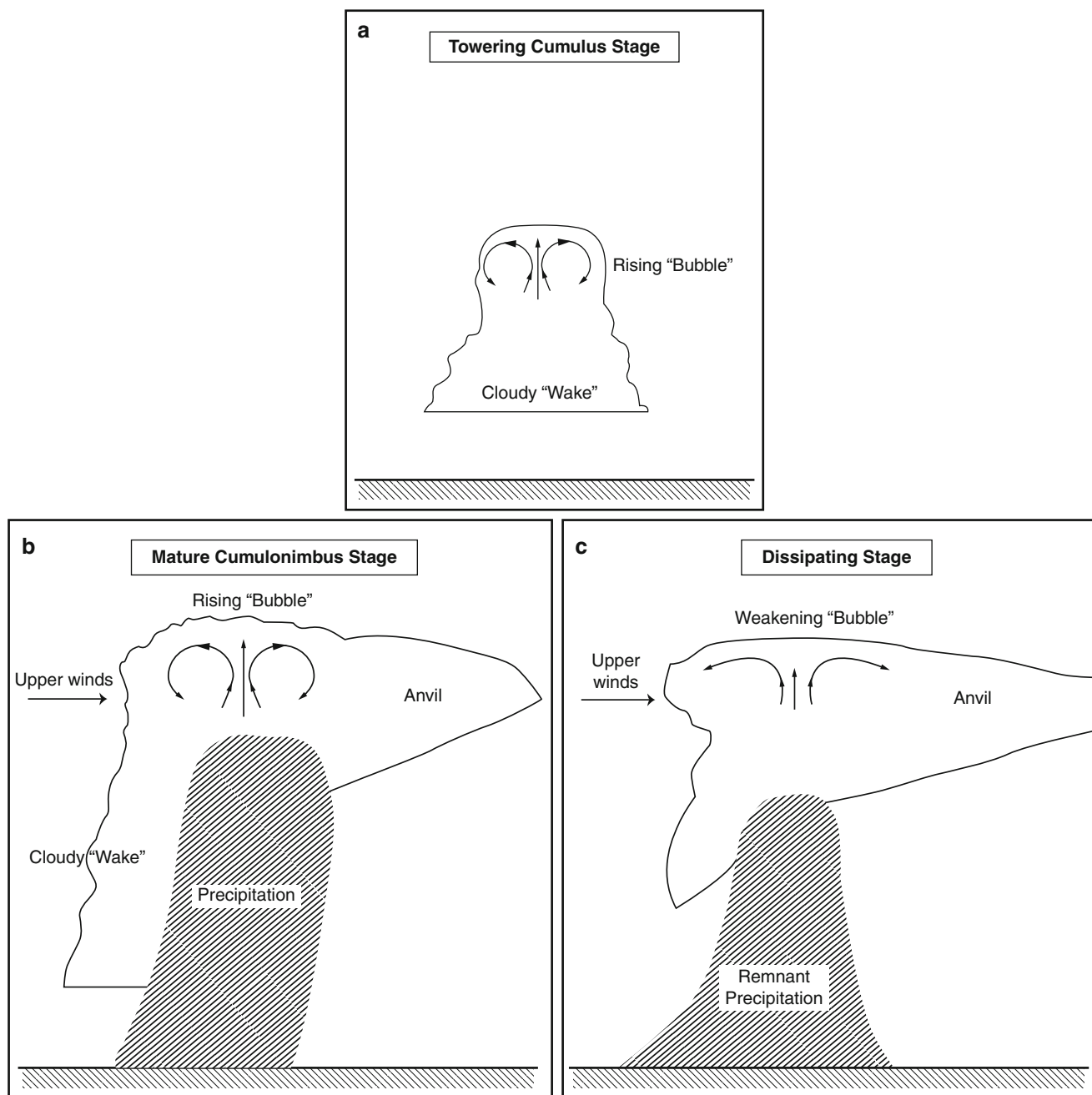
Any deep convective storm is made up of one or more of what are called *cells*. Each cell begins as a current of ascending air. These are discrete elements in space because of a property of fluids known as *mass continuity* – when air in one place arises, air from its surroundings must descend to take its place. The understanding of the cellular nature of convective storms was the result of a field observational campaign in the years following World War II, known as the Thunderstorm Project (Byers and Braham, 1949). During those field experiments, thunderstorms were observed by radars (which were then a relatively new technology) and penetrated by aircraft at multiple levels. The life cycle of an individual cell was determined from these observations to be about 20–40 min, corresponding roughly to the time it takes air to ascend from near the surface to the top of the storm. This life cycle is illustrated schematically in [Figure 2](#). Most thunderstorms

encompass multiple cells, and so can have lifetimes exceeding that of any individual cell.

Convective storms have updrafts that lift warm, moist air into the upper parts of the troposphere, usually resulting in precipitation. That ascent removes the excess latent and sensible heat at low levels, depositing it in the upper troposphere. Conversely, the storm's downdrafts take potentially cold air from middle and upper tropospheric levels and cause it to descend – which replaces warm air at low levels with that potentially cold air. Thus, updrafts and downdrafts both serve to alleviate the condition of excess heat at low levels. Once the instability that gave rise to the storm is reduced far enough by overturning within the deep convective storms, those storms will cease to exist. Convective storms can be seen as a response to a disturbance to equilibrium that results from solar heating (or from any anomalously warm underlying surface, such as warm surface currents in the oceans).

Severe convective storms

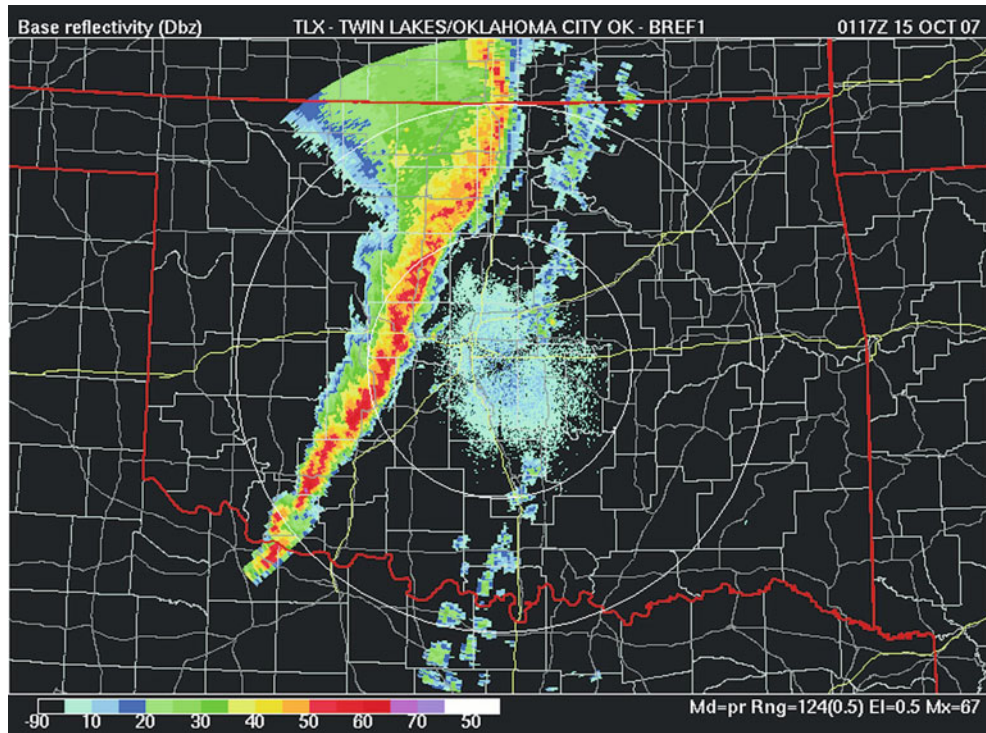
To this point, the processes described are common to both severe and nonsevere deep convective storms. What is it that makes a storm *severe*? To some extent, because there are thresholds for defining what constitutes a severe weather event involving hail or strong winds (Galway, 1989), the definitions are rather arbitrary. Is a storm producing a 1.8 cm hailstone qualitatively different from



Severe Storms, Figure 2 Schematic depiction of the life cycle of a single thunderstorm cell – (a) towering cumulus stage, (b) mature cumulonimbus stage, and (c) dissipating stage (From Doswell III (2001)).

a storm producing a 1.9 cm hailstone? Although any quantitative threshold between categories is arbitrary, as a given event passes a particular threshold, there is an increasing likelihood that important damage will result from that event. The impact of a storm also depends strongly on what it is affecting – a severe storm over open grassland is unlikely to do much damage to that grassland unless the

events become extreme. But within inhabited areas, damage with important consequences to humans could result from convective storm events even slightly below a particular threshold. Ultimately, any deep convective storm producing lightning has the potential to do damage and cause casualties, but lightning does not meet any of the existing formal criteria for a nominally severe storm.



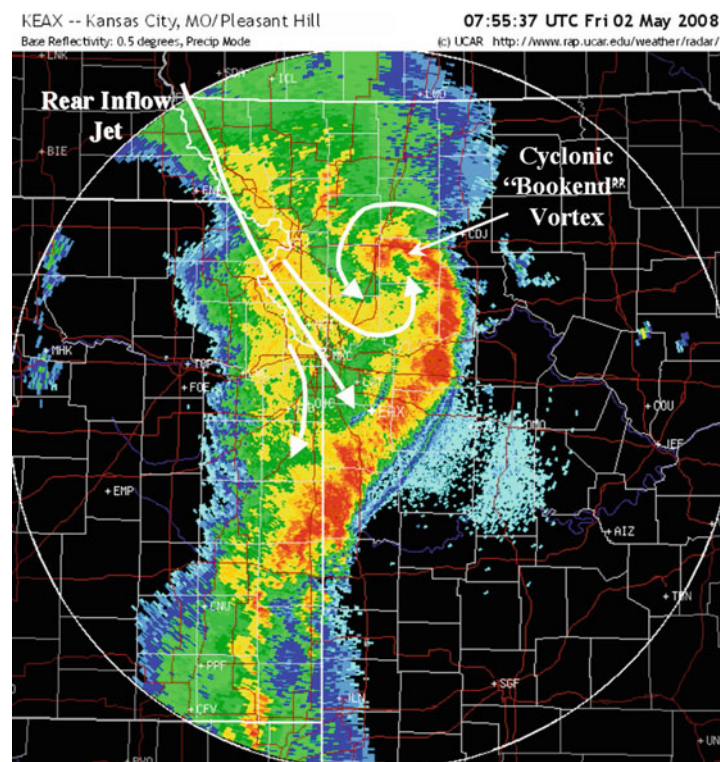
Severe Storms, Figure 3 An example of an event with a line of severe thunderstorms as seen by radar.

Severe weather is strongly associated with the degree to which the convective storms are organized. There are two basic modes by which convective storms become organized: into more or less continuous *lines* of individual convective cells, or as *discrete, relatively isolated* systems of cells. The ultimate form of discrete organization is the so-called supercell storm, which is characterized by a systematic rotation of the entire storm system through most of its depth. As will be discussed below, the main factor governing the degree of organization is the variation of winds with height.

The weakest degree of organization to a severe storm is the so-called pulse severe storm. Such an event is characterized by a very brief period (only a few minutes duration) of marginally severe weather, followed by rapid dissipation of the storm. Another pulse severe storm might develop in the vicinity, but it would be a distinctly different storm. Marginally severe hail and/or surface winds are the predominant events associated with pulse-type storms. At times, even weak updrafts can produce isolated strong downdrafts. The processes that control updraft strength are different from those controlling downdraft strength, so there is no physical reason to assume a high correlation between updraft and downdraft intensity in any given storm. Storms with weak updrafts might not even produce lightning but may result in small but powerful downdrafts called *microbursts* that can have

important human consequences, especially for aircraft taking off or landing (Fujita and Caracena, 1977).

When multiple storms develop in relatively close proximity, it is common for their surface outflows to merge, forming large pools of relatively cold air spreading outward along. The gust front at the leading edge of these merged outflows often serves to initiate new convective cells that then contribute their outflows to the cold pools, which then initiate another series of new convective cells, and so on. Furthermore, many storms are initiated along lines of ascent (typically, frontal boundaries), which imposes a linear structure on the resulting convective system. This linear organization (as in Figure 3) is very efficient at overturning large regions of unstable conditions. Strong surface winds arise when the downdrafts of such storms become intense, sometimes producing so-called bow echo morphology to the convective line (Figure 4). Severe hail is also common with such storms. The size of organized convective lines can vary from 20- to 200 km long, or even larger. At times, lines of storms (often referred to as *squall lines*) can produce swaths of damaging winds over regions encompassing 2,000 km² or larger – these events, meeting certain arbitrary criteria, are called *derechos* (Johns and Hirt, 1987). Derechos are an extreme form of convective wind event and if they happen to hit a populated area, the result can be devastating. Some derechos incorporate embedded supercell storms,



Severe Storms, Figure 4 Example of a bow echo as seen on radar, with white lines and arrows indicating the airflow. See text for discussion.

which often are responsible for the most intense winds and the largest hail. Tornadoes can occur with such linear structures, although not typically of the highest intensities unless they occur in association with embedded supercells.

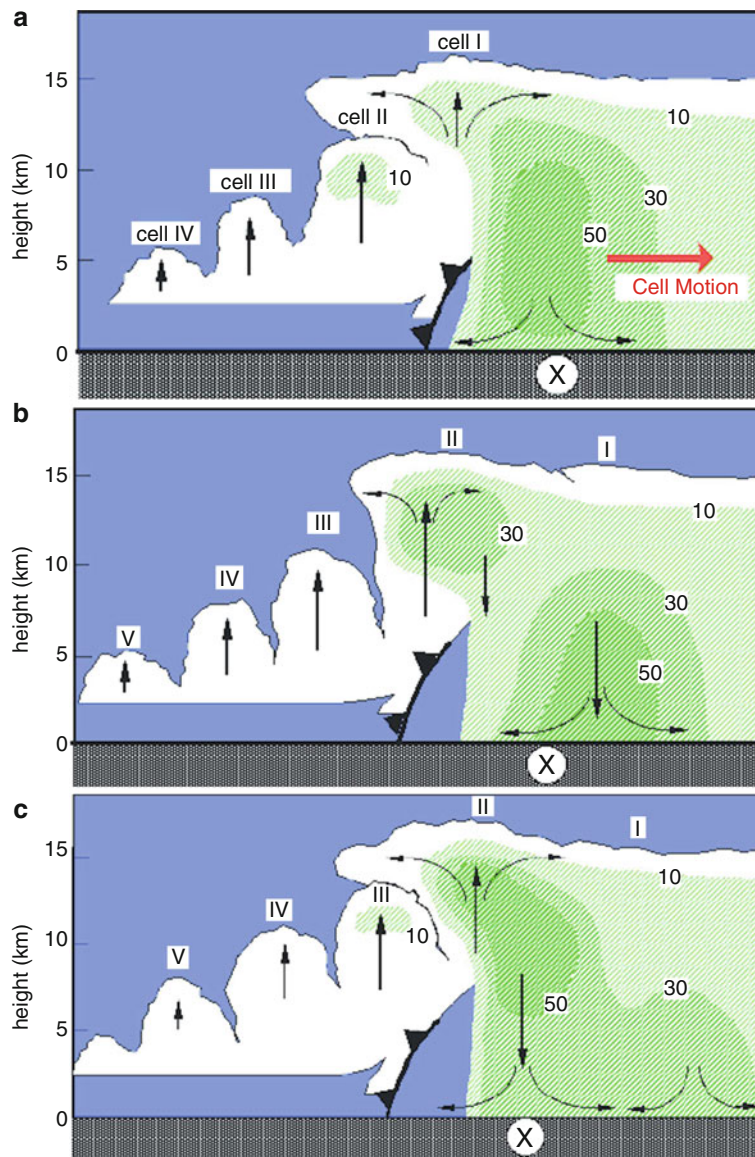
Another form of organization arises when the environment favors the development of new convective cells in roughly the same storm-relative position, resulting in an *isolated* multicellular storm (Marwitz, 1972b). Although it appears to be a single storm on radar, it really consists of a series of individual cells that are forming, maturing, and dissipating in an organized way. In effect, each cell is moving through the storm system as a whole. Such storms can produce strong winds and large hail, and occasional tornadoes. The severity of such events in a few cases can be quite high, but typically, the severe weather is of marginal to moderate intensity. However, such storms are a primary mode for the occurrence of heavy rainfall, when the storm system moves slowly while the individual cells move through the system (Figure 5).

The supercell (Browning, 1964; Marwitz, 1972a; Doswell III and Burgess, 1993) is the most organized type of severe storm, and almost all supercells produce some form of severe weather. The most extreme forms of severe weather are associated with supercells (Figure 6). If hailstone diameters reach 5 cm or more, the storm that produced them is likely to be a supercell. The most violent

tornadoes are almost exclusively confined to supercell storms (Davies-Jones et al., 2001). Even supercells possess multicellular characteristics (Foote and Frank, 1983), with an internal evolution cycle governed by the rotation process embedded within them, known as the *mesocyclone*. Mesocyclones evolve with time as the individual convective cells move through the storm, resulting in the cyclic production of severe weather, sometimes including families of tornadoes. In a few cases, supercells persist for many hours, producing swaths of severe weather that can be 200 km long or even longer.

Ingredients for large hail

The processes that result in large hail (Browning, 1977; Knight and Knight, 2001) are not entirely understood, but some conditions clearly are necessary. A strong updraft is required to hold the growing hailstone aloft long enough to attain their size – for a giant hailstone (say, 6 cm or larger), these updrafts must be about 50 m s^{-1} or more. Supercells are unique in that they occur in environments where the winds increase rapidly with height (i.e., strong *vertical wind shear*). When an updraft develops in an environment with strong vertical wind shear, the updraft can interact with that shear to produce a *perturbation* pressure gradient force that can augment the effect of positive buoyancy. That augmentation to the updraft's

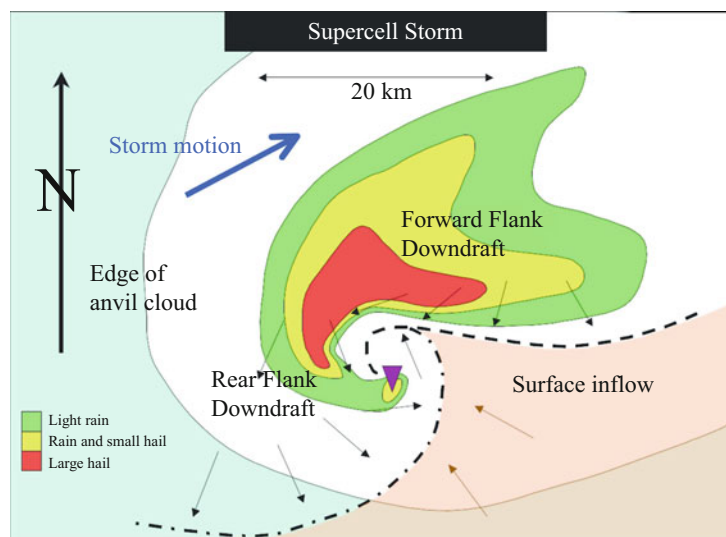


Severe Storms, Figure 5 Schematic illustrating the passage of individual convective cells through an isolated multicellular storm. Individual cells, labeled with Roman numerals I–V, form on the left and move to the right. The outflow's gust front is denoted by the black line with triangles. Precipitation intensity is indicated by the green shading. The point indicated by the circled "X" is a fixed point on the ground, which experiences repeated episodes of heavy rain coming from a sequence of cells moving through the storm system. The time required for the evolution in (a–c) is on the order of 20–30 min (Adapted from Figure 7 in Doswell et al. 1996).

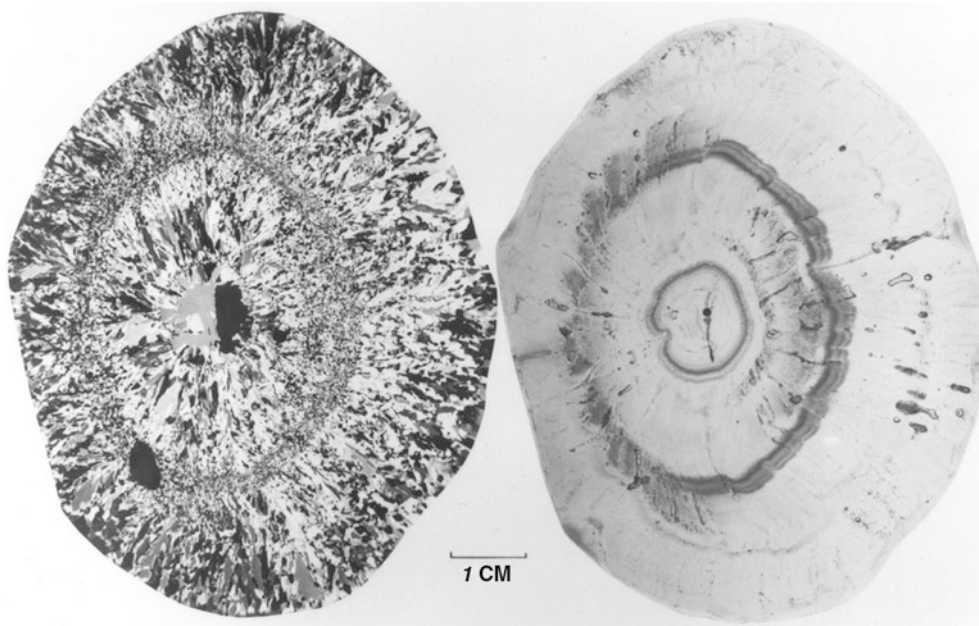
acceleration can result in a contribution as large as that from buoyancy alone. Thus, supercells generally have the strongest updrafts of any deep convective storms, which enhances their potential for giant hailstones. Not all supercells produce large hail, however. This implies that there are other ingredients for large hail. One likely candidate is the presence of large amounts of supercooled water in the updraft.

Liquid water condensing in clouds is typically very pure water and if carried above the level in the storm where the internal cloud temperature is 0°C , such pure

water may not freeze immediately. Very pure water's temperature can be lowered to about -40°C , the so-called homogeneous nucleation temperature, before it freezes, unless there is a *freezing nucleus* present. One type of freezing nucleus is water already frozen: a frozen raindrop or a snow particle, for example. In strong updrafts with copious condensed water within them, the presence of supercooled water is likely. When a supercooled water droplet makes contact with ice, it freezes quite rapidly, so hailstones grow by a process of accreting supercooled liquid water – called *wet growth*. Hailstones also can grow



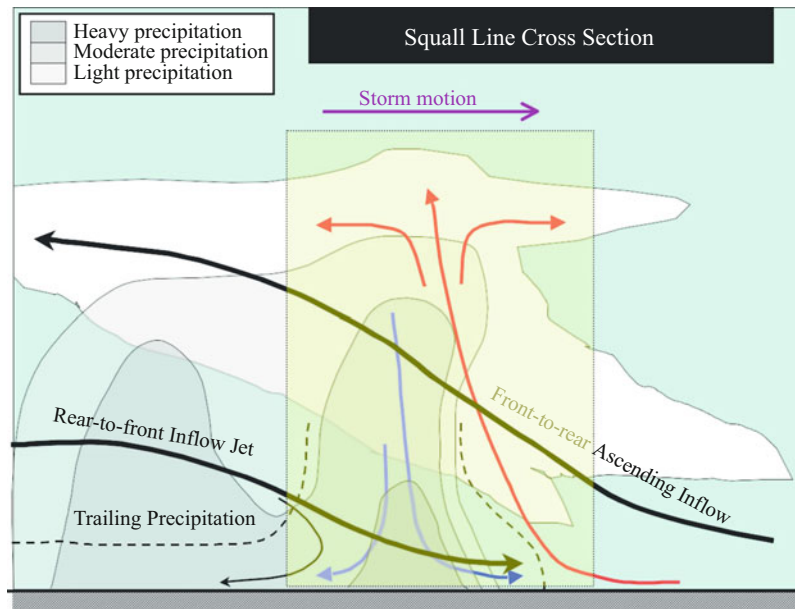
Severe Storms, Figure 6 Schematic view of a supercell storm as it might appear on radar, showing the location of important features associated with such storms, with small *black arrows* indicating the low-level airflow. The *dashed black line* is the gust front associated with the forward-flank downdraft, while the *dash-dot black line* is that associated with rear-flank downdraft. The location of a mesocyclonic tornado is shown by the *purple inverted triangle*.



Severe Storms, Figure 7 A cross section of a large hailstone, showing the “growth rings” produced by alternating wet and dry growth of the hail. On the *left*, the view is through crossed polarizers, showing the internal crystalline grain structure. On the *right*, through ordinary transmitted light – it is evident that cloudy ice has smaller grain sizes (Photograph courtesy C. Knight (used with permission)).

by *dry growth*, which is the conversion of water vapor directly into solid form (sometimes called *riming*). As the growing hailstone alternates between periods of wet and dry growth, its internal structure exhibits layering

between nearly transparent ice from wet growth and translucent ice from dry growth (Figure 7). The notion of hailstones acquiring this layered structure from repeated ascent to high levels and descent to low levels in the storm



Severe Storms, Figure 8 Schematic cross section of the airflow in a typical mesoscale convective system (MCS). The deep convective cells are along the leading edge of the system, with updrafts depicted in the *red lines* and downdrafts shown in *blue*, but airflow also goes *in between* the cells along an ascending path from front to rear (*heavy black line*). In some systems, a rear-to-front descending inflow jet (*heavy black line*) develops, entering the storm at middle levels and can result in a bow echo configuration for the line of storms (as in [Figure 4](#)). The box outlined with *dotted lines* and colored *light yellow* contains the active deep convective part of the system (cf. the *red colors* indicating heavy rain and possibly hail in [Figures 3 and 4](#)).

has now largely been discredited (Knight and Knight, 2001). Minor height fluctuations and movement within the storm can cause the growth mode to alternate between wet and dry without requiring cycles of substantial ascent and descent through the storm.

In order to reach the ground as a large hailstone, once the hailstone grows large enough to begin its descent from the growth zone, it should not experience an extended period of melting once it falls below the melting level (0°C). It turns out that a relatively dry environment inhibits melting during descent, as well. There likely are microphysical factors that promote or inhibit the development of large hail, but these remain unknown at present.

Ingredients for strong winds

As discussed earlier, two primary factors promote strong downdrafts, with their associated strong surface winds: negative buoyancy produced by evaporation (or by direct conversion of ice to vapor, known as *sublimation*), and precipitation loading (Wakimoto, 2001). It is possible that another contribution to strong outflow is perturbation pressure gradient forces. At present, the extent to which such pressure forces are involved is not known comprehensively. The interaction between a downdraft and the surface produces strong vertical deceleration of the updraft, which results in an outward-directed perturbation pressure gradient force that drives the outflow once the downdraft descends far enough to feel the effects of the surface. There may also be an interaction between

the downdraft and the environmental vertical wind shear, but the importance of that is yet to be shown.

In supercell storms, the mesocyclone provides an augmentation of the pressure gradient force that causes a swirling airflow. This results in an increase of the magnitude not only of the downdraft's outflow, but also the updraft's inflow. Thus, some supercells may have narrow channels of potentially damaging winds in their inflows, as well as within their outflows.

In large systems of deep convection, called *mesoscale convective systems* (or MCSs) (Fritsch and Forbes, 2001), a relatively narrow band of inflow can develop a few kilometers above the surface and enter the storm system from the rear (illustrated schematically in [Figure 8](#)). As precipitation falls into this rear inflow, it is cooled by evaporation and can descend to the surface where it can augment the convective storm-scale downdrafts driving the gust front. This process may explain the bowed-shaped morphology of many bow echoes. Because of the relatively narrow jet of strong winds into the rear of an MCS, strong horizontal wind shear on its flanks produces counterrotating vortices on the ends of bow echoes, called "bookend" vortices (cf. [Figure 4](#)). Also shown in [Figure 8](#) is the inflow that passes *between* deep convective cells forming on the leading edge of the outflow and ascends gradually toward the rear of the storm. This carries precipitation toward the trailing side of the storm, where it falls out as moderate rainfall. Generally speaking, the *cyclonic* vortex (see the next section) can be augmented by the

Coriolis effect (an apparent acceleration of the airflow due to the Earth's rotation) over the life cycle of an MCS, whereas the *anticyclonic* vortex is inhibited by the Coriolis effect.

Ingredients for tornadoes

Not all tornadoes arise the same way, so the task of describing the factors that promote the development of tornadoes (tornadogenesis) is somewhat challenging. It has been mentioned that tornadoes occasionally form from isolated non-supercell storms (Wakimoto and Wilson, 1989), but even within supercells, it appears that tornadoes can develop in different ways. Hence, the distinction made here is between tornadoes that occur embedded within mesocyclones and those that occur in the absence of, or some distance away from, a mesocyclone (Davies-Jones et al., 2001). The former are mesocyclonic tornadoes, the latter are non-mesocyclonic tornadoes (Brady and Szoke, 1989).

For discussing tornadoes, it is useful to define a quantity associated with fluid motion called *vorticity*. Vorticity is a vector, denoted by $\mathbf{\Omega}$, defined as the curl of the velocity field:

$$\mathbf{\Omega} \equiv \nabla \times \mathbf{V},$$

where the velocity vector $\mathbf{V}$ has components (u, v, w) in the eastward, northward, and upward vertical directions, respectively in a rectangular Cartesian coordinate system (x, y, z) , with unit vectors $(\mathbf{i}, \mathbf{j}, \mathbf{k})$. Vorticity follows a "right-hand rule" convention. If the direction of the vorticity vector is parallel to and in the same direction as the thumb on the right hand, the slightly curled fingers are aligned with the sense of rotation (counterclockwise) and the vorticity's magnitude is positive. Negative vorticity is antiparallel to the thumb of the right hand. Vorticity has three components:

$$\mathbf{\Omega} = \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right) \mathbf{i} + \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \right) \mathbf{j} + \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) \mathbf{k} \equiv (\xi, \eta, \zeta).$$

The vorticity in a tornado is primarily about a vertical axis (the ζ -component of the vector vorticity). For a tornado, the peak value of the vertical vorticity is on the order of 1.0 s^{-1} , whereas for a mesocyclone, the peak value is of order $1.0 \times 10^{-2} \text{ s}^{-1}$, two orders of magnitude less (Doswell III and Burgess, 1993).

When speaking of cyclones and anticyclones, in the Northern Hemisphere, the vorticity in a cyclone is positive (counterclockwise) and negative (clockwise) in an anticyclone. In the Southern Hemisphere, the vorticity of a cyclone is negative and positive in an anticyclone.

The issue of tornadogenesis concerns the mechanism(s) by which the vertical vorticity becomes so large, especially near the surface (Davies-Jones et al., 2001). At one point in the study of tornadoes, it was felt that it was simply a matter of amplification of the mesocyclone's vorticity by conservation of angular momentum within the updraft. However, observations using Doppler radar have

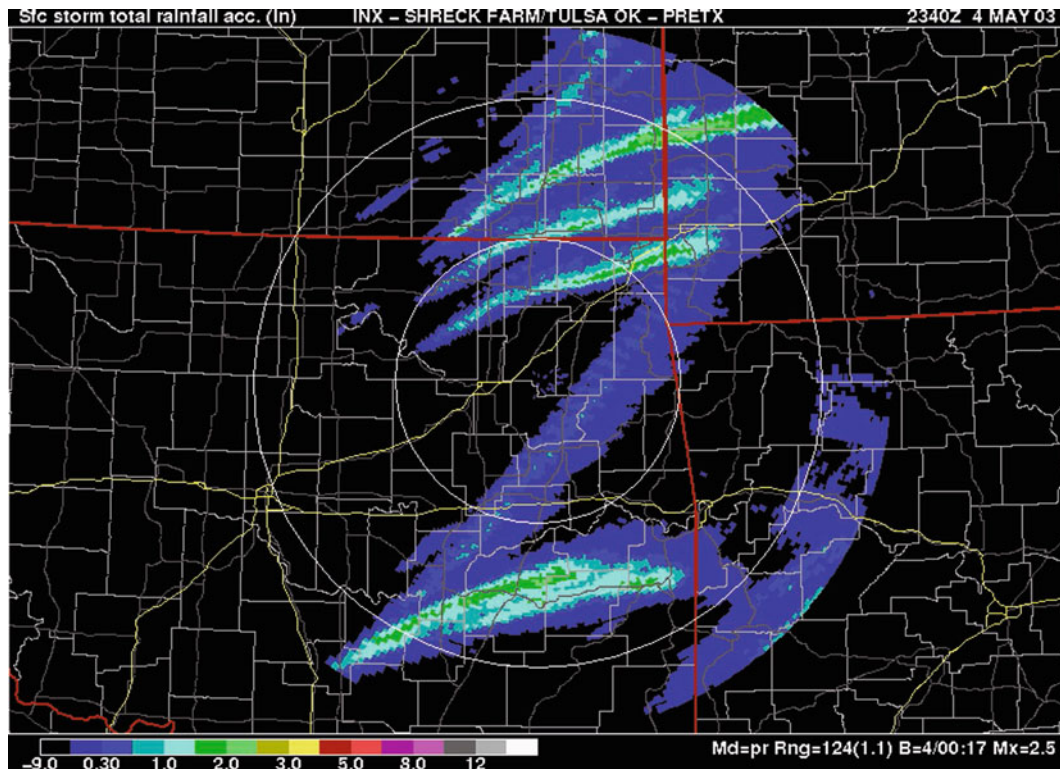
made it clear that this simple idea is not so easily applied. Mesocyclonic tornadoes typically form near the interface between updraft and downdraft (Lemon and Doswell, 1979), whereas if they were associated only with simple conservation of angular momentum, it would be expected they would form near the center of the updraft. While there can be no doubt that conservation of angular momentum is involved, the specific internal storm and environmental processes that promote tornadogenesis in a few storms and not in most others remain unknown. Only around 20 % (or perhaps less) of supercells produce tornadoes (the percentage for other storm organization levels is much lower than for supercells), so the presence of mesocyclonic vorticity is not by itself sufficient to explain tornadogenesis. Further, the fact that some convective storms without mesocyclones can produce tornadoes means that other, presently unknown factors must be involved.

The primary mechanism for producing supercells is reasonably well known. When an environment is characterized by strong vertical wind shear through a deep layer (6 km or more above the surface), with magnitudes of order $3 \times 10^{-3} \text{ s}^{-1}$ or more, this implies ambient vorticity about a horizontal axis. The two components of this horizontal vorticity vector are $\mathbf{\Omega}_h = (\xi, \eta)$ and for storm environments, that horizontal vorticity is dominated by the two terms involving the vertical wind shear, so that on large scales,

$$\xi \approx -\frac{\partial v}{\partial z}, \quad \eta \approx \frac{\partial u}{\partial z},$$

because the vertical component of the airflow (w) in a convective storm's environment is at least two orders of magnitude smaller than the horizontal components. This means that the horizontal component of the environmental vector vorticity is very nearly perpendicular to the vertical wind shear vector, $\mathbf{S}$, given by $\mathbf{S} \equiv \partial \mathbf{V}_h / \partial z$, where $\mathbf{V}_h = (u, v)$.

The horizontal vorticity also can be decomposed into components along and perpendicular to the airflow, known as *streamwise* and *crosswise* vorticity, respectively. Davies-Jones et al. (2001) demonstrate that an updraft causes crosswise vorticity to be tilted into the vertical to form a pair of counterrotating vortices: one cyclonic and the other anticyclonic. When the vorticity vector is parallel to the airflow, that vorticity is either streamwise or antistreamwise. Supercell environments favor streamwise vorticity through a deep layer, which is what we generally observe, although in some circumstances, supercells can occur with a large component of crosswise vorticity, as well. In such instances, a common observation is the production of counterrotating supercells, the cyclonic one moving to the right of and slower than the mean wind through a deep layer, and the anticyclonic one moving to left of and faster than the mean wind (Figure 9), in the Northern Hemisphere. As usual, things work the opposite way in the Southern



Severe Storms, Figure 9 Tracks of left- and right-moving supercells, as shown by the accumulated rainfall amounts over the life of the echoes. The fast-moving left movers do not accumulate much rainfall, but the relatively slow-moving right movers produce heavier rainfall amounts and so show brighter colors along their tracks.

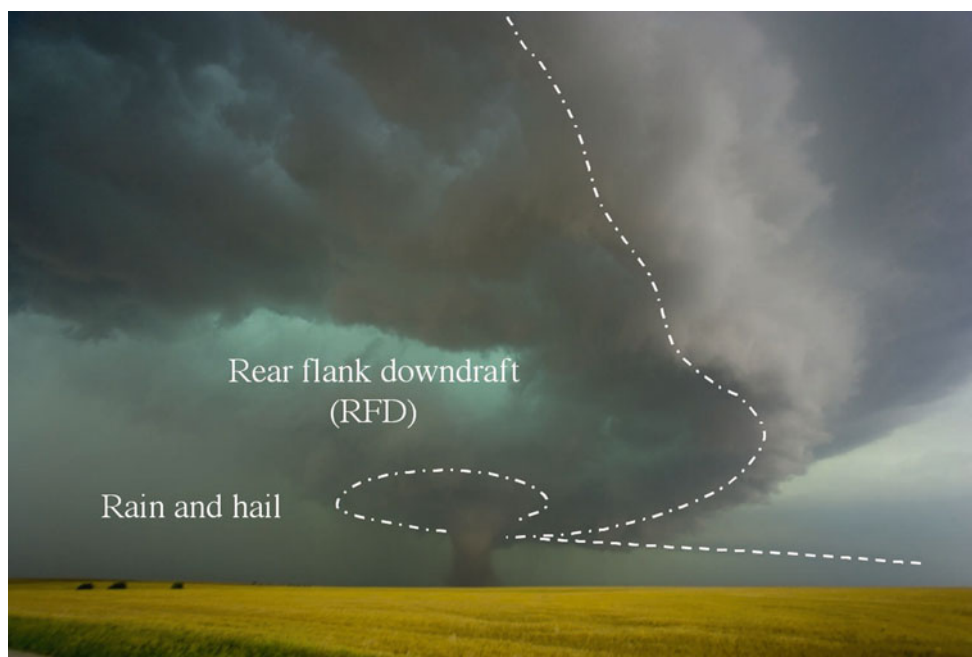
Hemisphere – the cyclonic storm moves to the left and slower than the mean wind, whereas the anticyclonic supercell moves to the right and faster than the mean wind.

Although the physical understanding needed to define the specific ingredients for tornadogenesis is lacking, empirical results suggest two factors that seem to be important (Doswell and Evans, 2003; Thompson et al., 2003). The first is the presence of strong vertical wind shear in the lowest kilometer of the atmosphere – vertical wind shear values approaching $1.0 \times 10^{-2} \text{ s}^{-1}$ seem to be common in tornadic storm environments, be they supercells or not. The second is high relative humidity near the surface, which seems to be related to yet another empirical finding about tornadic supercells (Figure 10): Their downdrafts tend to be relatively warm and may even retain some buoyancy if lifted far enough. A cold, very stable outflow from a storm reduces the likelihood of tornadoes and such outflows are promoted by low relative humidity in the near-surface air. Again, there is as yet no comprehensive physical explanation for these empirical results.

Much less is known about tornadoes produced by non-supercell storms. It appears that in at least some cases, the existence of preexisting maxima of vertical vorticity on a scale larger than that of the convective storm (called *misocyclones*) means that it may be simple

amplification of that vorticity by developing updrafts through conservation of angular momentum which can explain this type of tornado (Brady and Szoke, 1989). Non-mesocyclonic tornadoes have not been studied very extensively, so much continues to be unexplained. It should be noted that waterspouts are simply tornadoes over the water, and some of them also seem to be associated with supercell storms and others with non-supercellular deep convection. There are areas around the world that produce frequent waterspouts from primarily non-supercellular storms. Many of these waterspouts arise in deep convection that has very little or no lightning activity, at least prior to waterspout development.

An important scientific question relating to tornadoes is their role in the atmosphere. Most meteorological processes have a well-defined purpose; for example, deep, moist convection is a response to the excessive sensible and latent heat at low levels and convective storms result in alleviating the instability that gives them birth. Tornadoes almost certainly have some role to play, because the alternative would be that they are just random events associated with deep convective storms, which seems unlikely. Our inability to predict which storms will produce tornadoes is evidently related to this currently unanswered question. The atmosphere is an example of



Severe Storms, Figure 10 A tornadic supercell, with annotations, showing the location of the rear-flank (*white dash-dot line*) and forward-flank (*white dashed line*) gust fronts (cf. Figure 6) at cloud base level. This example is looking roughly toward the northeast at a storm like that illustrated in Figure 6. The rear-flank downdraft is creating the *turquoise-colored* “hole” in the clouds seen above and toward the left of the tornado (Original photograph © 2005 C. Doswell, annotated version © 2008 C. Doswell).

a nonlinear dynamical system (Lorenz, 1993) and it is known that such systems can be difficult to predict even when the underlying dynamical processes are understood perfectly. Gaps in that understanding make it even more challenging to anticipate such events.

Convective storm observations

Because convective storms are small relative to the sparse conventional in situ observing systems, a great deal of what we know about them has been derived from remote sensing – primarily radar. The development of Doppler radar for probing convective storms began in the late 1960s and has become the basis for our existing operational severe weather warning system. Recently, the implementation of polarimetric capability for the operational Doppler radars has begun, which will give additional information for observing the microphysical structure within convective storms.

Another important observing capability has been spacecraft with cameras and other active and passive remote sensors. The distribution of deep convective storms in space and time via satellite observations has been another important component in scientific understanding. Further, it is possible to obtain critical information about the distribution of environment variables relevant to deep convection using a number of different spacecraft-borne observing systems, including lightning detection capability. In the past two decades, both spacecraft and ground-based remote sensing tools have

been developed that offer considerable promise to revolutionize our ability to sample the global environment associated with deep convection.

Summary

The definition of severe weather as used here is based on arbitrary criteria for hail and strong winds. The storms that produce large hail are clearly those that include strong updrafts, and since supercells develop the strongest vertical motions of any convective storm, the largest hail is produced primarily by supercells. Not all supercells result in large hail; however, predicting the occurrence of large hail is still somewhat problematic. It is likely that the detailed microphysical processes within convective storms are important for the production of hail but much remains to be learned about microphysical processes in deep convective storms. Polarimetric radars, which are just now on the verge of becoming operational, may eventually help with understanding hail production. However, a challenge for using polarimetric radars (and other remote sensing tools) to infer microphysical properties in the atmosphere is that in situ validation of the interpretation of the observations in terms of the microphysics is needed. Are the remote sensing observations being interpreted properly? This continues to be a vexing issue.

Strong winds are arguably the easiest type of severe weather event to understand. They are almost entirely associated with strong downdrafts, the physics of which

is reasonably well understood. It is not always possible to know in detail what a particular storm is likely to do, but the general conditions for producing a strong surface wind are relatively easy to anticipate. One challenge is to know when Doppler radar-observed strong winds are actually making their way to the surface. When a shallow stable layer is present at the surface, the strong winds just above may or may not be observed at levels near the surface where they can do damage. Radars have a “horizon” problem associated with the curvature of the Earth—at its lowest scanning angle, it observes events only above a height that increases with distance. Hence, Doppler radars generally cannot confirm or deny that strong winds observed above the surface are actually reaching down far enough to do damage.

At this time, it is not clear just what role tornadoes are playing that other physical processes apparently cannot accomplish. Whatever that purpose might be, the need for it is relatively infrequent—tornadoes are rare events in a global sense. The challenge posed by tornadogenesis is connected to the reason for their existence in the first place—if their role in the atmosphere can be understood definitively, the mechanism(s) by which they are created should become more clear than at present. A major concern for hypothesized tornadogenesis mechanisms is that they need to be able to account for non-mesocyclonic tornadoes, as well. Most of our observations of tornadoes are for mesocyclonic events associated with supercells. Research in the near future is likely to remain concentrated on trying to explain mesocyclonic tornadoes, seeking to understand the physical processes that explain the empirical associations with low-level vertical wind shear and high relative humidities near the surface, but the broader topic is likely to remain problematic for some time, as non-mesocyclonic tornadoes (including waterspouts) seem much harder to predict. New technologies for observing the airflow and thermodynamic characteristics in the lowest kilometer of the atmosphere will need to be developed if much progress is to be made.

Another remaining challenge is the development of an accurate understanding of the true distribution of severe weather (Brooks et al., 2003; Doswell et al., 2005). Even in the United States, which has the best infrastructure for documenting severe weather in the world, hail and strong winds are reported as point events, whereas the reality is that both winds and hail occur in swaths. Thus, the reports of severe weather at points are likely to be providing only a biased depiction of the frequency and intensity of severe weather. Such reports depend on having a knowledgeable observer at the right place and time to observe the most intense events, an unlikely scenario. Remote sensing systems (mostly radar) do not make direct observations of severe weather events but may offer the best hope for an accurate spatiotemporal distribution of severe storms.

Tornadoes are assigned path lengths and widths, but at the moment, the occurrence of tornadoes (and other forms of severe weather) in sparsely populated regions often goes unreported. Further, there simply are not enough

available resources to devote to a careful study of each reported tornado event and the intensity determination is based almost entirely on the degree and type of damage done. When a tornado occurs in sparsely populated regions of grassland, with little or nothing to damage, no way exists at present to determine the tornado’s wind speeds. If reported at all, such tornadoes typically are given a “default” rating as weak events, which can be substantially wrong. Moreover, even when careful damage surveys are done, there is no simple relationship between wind speed and damage. Doing a damage survey to estimate wind speeds requires considerable experience and knowledge, and this is not generally available. Until some way to estimate tornado intensity objectively and remotely is developed, comparable to the Richter scale in estimating earthquake intensity, our knowledge of the distribution and intensity of tornadoes will remain uncertain.

Bibliography

- Brady, R. H., and Szoke, E. J., 1989. A case study of nonmesocyclone tornado development in Northeast Colorado: similarities to waterspout formation. *Monthly Weather Review*, **117**, 843–856.
- Brooks, H. E., Doswell, C. A., III, and Kay, M. P., 2003. Climatological estimates of local daily tornado probability for the United States. *Weather and Forecasting*, **18**, 626–640.
- Browning, K. A., 1964. Airflow and precipitation trajectories within severe local storms which travel to the right of the winds. *Journal of the Atmospheric Sciences*, **21**, 634–639.
- Browning, K. A., 1977. The structure and mechanism of Hailstorms. In Knight, C. A. (ed.), *Hail: A Review of Hail Science and Hail Suppression*. Boston: American Meteorological Society. Meteorological Monographs, Vol. 38, pp. 1–39.
- Byers, H. R., and Braham, R. R., Jr., 1949. *The Thunderstorm*. Washington, DC: U.S. Government Printing Office, p. 287. [out of print].
- Davies-Jones, R., Trapp, R. J., and Bluestein, H. B., 2001. Tornadoes and tornadic storms. In Doswell, C. A. (ed.), *Severe Convective Storms*. Boston: American Meteorological Society. Meteorological Monographs, Vol. 28, No. 50, pp. 167–221.
- Doswell, C. A. III, H. E. Brooks, and R. A. Maddox, 1996. Flash flood forecasting: An ingredients-based methodology. *Weather and Forecasting*, **11**, 560–580.
- Doswell, C. A. III, 2001. Severe convective storms – an overview. In Doswell, C. A. III (ed.), *Severe Convective Storms*. Boston: American Meteorological Society. Meteorological Monographs, Vol. 28, No. 50, pp. 1–26.
- Doswell, C. A. III, and Burgess, D. W., 1993. Tornadoes and tornadic storms: a review of conceptual models. In Church, C., et al. (eds.), *The Tornado: Its structure, Dynamics, Prediction, and Hazards*. Washington DC: American Geophysical Union. Geophysical Monographs, No. 79, pp. 161–172.
- Doswell, C. A., III, and Evans, J. S., 2003. Proximity sounding analysis for derechos and supercells: an assessment of similarities and differences. *Atmospheric Research*, **67–68**, 117–133.
- Doswell, C. A., III, Brooks, H. E., and Kay, M. P., 2005. Climatological estimates of daily nontornadic severe thunderstorm probability for the United States. *Weather and Forecasting*, **20**, 577–595.
- Doswell, C. A. III, Brooks, H. E., and Dotzek, N., 2009. On the implementation of the enhanced Fujita scale in the USA. *Atmos. Res.*, **93**, 554–563.

- Foote, G. B., and Frank, H. W., 1983. Case study of a hailstorm in Colorado. Part III: airflow from triple-Doppler measurements. *Journal of the Atmospheric Sciences*, **40**, 686–707.
- Fritsch, J. M., and Forbes, G. S., 2001. Mesoscale convective systems. In Doswell, C. A. III (ed.), *Severe Convective Storms*. Boston: American Meteorological Society. Meteorological Monographs, Vol. 28, No. 50, pp. 323–357.
- Fujita, T. T., 1971. Proposed characterization of tornadoes and hurricanes by area and intensity. *Satellite and Mesometeorology Research Project Research Paper No. 91*. Chicago, IL: University of Chicago, p. 42.
- Fujita, T. T., and Caracena, F., 1977. An analysis of three weather-related aircraft accidents. *Bulletin of the American Meteorological Society*, **58**, 1164–1181.
- Galway, J. G., 1989. The evolution of severe thunderstorm criteria within the weather service. *Weather and Forecasting*, **4**, 585–592.
- Johns, R. H., and Hirt, W. D., 1987. Derechos: widespread convectively induced windstorms. *Weather and Forecasting*, **2**, 32–49.
- Knight, C. A., and Knight, N. C., 2001. Hailstorms. In Doswell, C. A. III (ed.), *Severe Convective Storms*. Boston: American Meteorological Society. Meteorological Monographs, Vol. 28, No. 50, pp. 223–254.
- Lamb, D., 2001. Rain production in convective storms. In Doswell, C. A. III (ed.), *Severe Convective Storms*. Boston: American Meteorological Society. Meteorological Monographs, Vol. 28, No. 50, pp. 299–321.
- Lemon, L. R., and Doswell, C. A., III, 1979. Severe thunderstorm evolution and mesocyclone structure as related to tornadogenesis. *Monthly Weather Review*, **107**, 1184–1197.
- Lorenz, E. N., 1993. *The Essence of Chaos*. Seattle: University of Washington, p. 227.
- Ludlam, F. H., 1963. Severe local storms: a review. In Atlas, D. (ed), *Severe Local Storms*. Boston: American Meteorological Society. Meteorological Monographs, Vol. 5, No. 27, pp. 1–30.
- Marwitz, J. D., 1972a. The structure and motion of severe hailstorms. Part I: supercell storms. *Journal of Applied Meteorology*, **11**, 166–179.
- Marwitz, J. D., 1972b. The structure and motion of severe hailstorms. Part II: multi-cell storms. *Journal of Applied Meteorology*, **11**, 180–188.
- Newton, C. W., 1963. Dynamics of severe convective storms. In Atlas, D. (ed.), *Severe Local Storms*. Boston: American Meteorological Society. Meteorological Monographs, Vol. 5, No. 27, pp. 33–58.
- Schultz, D. M., Schumacher, P. N., and Doswell, C. A., III, 2000. The intricacies of instabilities. *Monthly Weather Review*, **128**, 4143–4148.
- Thompson, R. L., Edwards, R., Hart, J. A., Elmore, K. L., and Markowski, P., 2003. Close proximity soundings within supercell environments obtained from the rapid update cycle. *Weather and Forecasting*, **18**, 1243–1261.
- Wakimoto, R., 2001. Convectively driven high wind events. In Doswell, C. A. III (ed), *Severe Convective Storms*. Boston: American Meteorological Society. Meteorological Monographs, Vol. 28, No. 50, pp. 255–298.
- Wakimoto, R. M., and Wilson, J. W., 1989. Non-supercell tornadoes. *Monthly Weather Review*, **117**, 1113–1140.

Cross-references

[Cloud Liquid Water](#)
[Earth Radiation Budget, Top-of-Atmosphere Radiation](#)
[Lightning](#)
[Observational Systems, Satellite](#)
[Ocean-Atmosphere Water Flux and Evaporation](#)

[Radiation, Electromagnetic](#)
[Rainfall](#)
[Water Vapor](#)
[Weather Prediction](#)

SNOWFALL

Ralf Bennartz

Atmospheric and Oceanic Sciences Department,
 University of Wisconsin-Madison, Madison, WI, USA

Definition

Snow. Low-density ice particles. The density is typically in the order of 0.1 g/cm^3 . Individual snowflakes can exhibit a wide variety of different forms. Other frozen particles include graupel or hail. Graupel particles are medium-density ice particles with a density of about 0.4 g/m^2 . Graupel is produced when ice particles fall through extensive layers of supercooled cloud liquid water. Hail particles are of a very high density of about 0.9 g/m^2 . Hail fall is associated with intensive convection and is formed via a series of melting and refreezing processes.

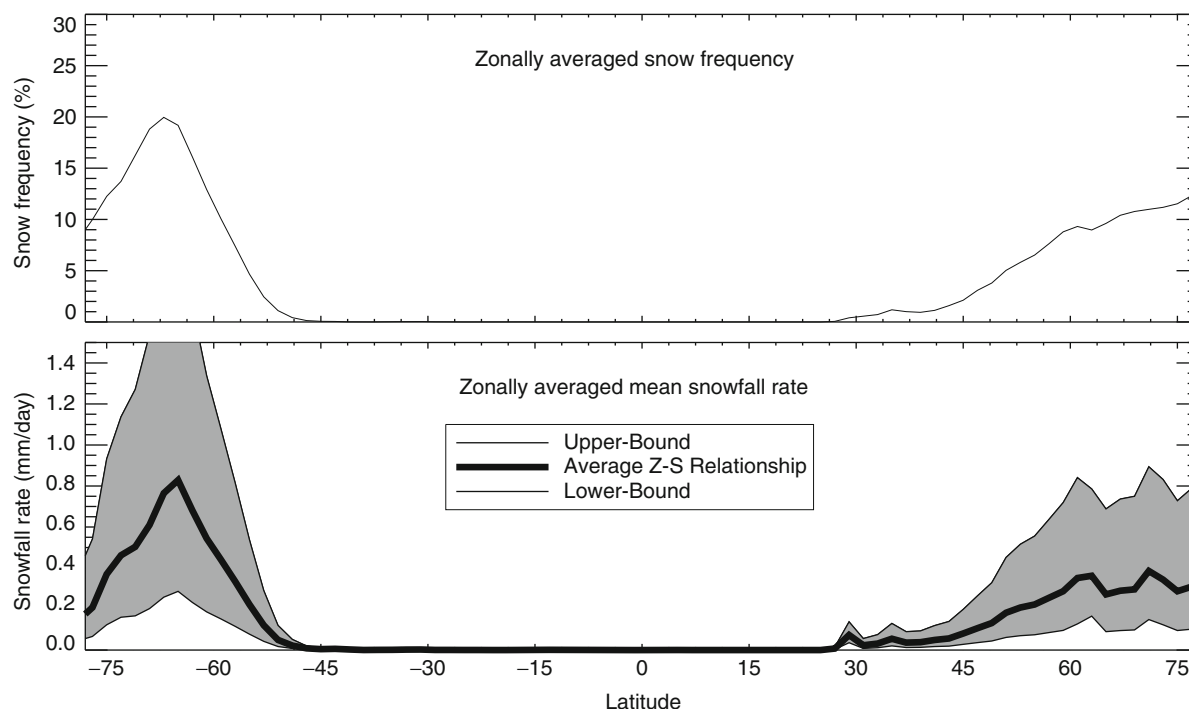
Melting layer. The melting layer is the layer in which falling hydrometeors transition from the ice into the liquid phase. The top of the melting layer coincides with the 0°C isothermal.

Bright band. The bright band is a band of enhanced radar reflectivity associated with melting precipitation particles. It is typically found in stratiform precipitation. The bright band roughly coincides with the melting layer.

Introduction

Frozen precipitation exists in a huge variety of different forms depending on weather situation as well as on cloud microphysical processes. Precipitation-sized ice particles can be broadly categorized into three classes: snow, hail, and graupel. All deep convective and frontal precipitation events involve the ice phase even if ice particles will melt before reaching the ground. Snowfall at the surface constitutes a large fraction of polar, boreal, and mid-latitude winter precipitation. In the northern hemisphere, winter snowfall is ubiquitous north of roughly 40°N to 50°N . In most mid- and high-latitude countries, snowstorms can have a major disrupting effect on traffic and the economy.

The atmospheric temperature profile of the atmosphere plays a key role in determining whether liquid, frozen, or mixed precipitation reaches the surface. In this regard, the problem of snowfall remote sensing can be separated into three different issues. Firstly, precipitation has to be detected. Secondly, the phase or type of precipitation at the surface has to be determined. Thirdly, the amount of precipitation has to be quantified. This entry is only concerned with snowfall at the surface and concentrates



Snowfall, Figure 1 Zonally averaged snowfall frequency and mean snowfall rate derived from 1 year of CloudSat radar observations. The uncertainty in the snowfall rate (gray-shaded area) is solely due to assumptions about ice habit (Figure from Hiley et al., 2010).

on the second and third issue. While there are various ground-based remote sensing techniques for snowfall, the main focus here is on global satellite-derived estimates. Subsequently, a distinction will be made between active (radar) and passive (microwave radiometer) remote sensing techniques.

Principles of snowfall remote sensing

Frozen versus liquid precipitation

The discrimination of frozen from liquid precipitation at the ground is clearly crucial, especially in high- and mid-latitude areas where snowfall occurrence varies widely on annual and interannual time scales. From a passive remote sensing standpoint, the determination of the phase of falling precipitation at the ground is an ill-posed problem, since passive measurements typically have broad vertical weighting functions. For instance, if the precipitation melts just a few meters above the surface, it will not have any measurable “liquid” signature that could be detected using a passive sensor. Passive instruments are therefore not directly sensitive to the phase of falling precipitation at any given level. Indirectly, the phase of precipitation at the surface can be determined with some accuracy by assessing the height of the freezing level. If the freezing level is known, the only additional information needed to assess precipitation phase at the surface is the distance it takes for the falling precipitation to melt. A reasonable value for this falling distance is in the order of

500–600 m for snowfall (e.g., Bennartz, 2007). As an example, the upper panel in Figure 1 shows the zonally averaged snowfall frequency based on temperature fields obtained from a numerical weather prediction model and precipitation identification using CloudSat’s 94 GHz Cloud Profiling Radar (Stephens et al., 2008).

Some active remote sensing techniques (radar) provide additional means to distinguish frozen from liquid precipitation. The occurrence of a bright band in radar observations is a sign for melting particles (Fabry and Zawadzki, 1995). Thus, if a bright band is identified sufficiently high above the surface, precipitation at the surface can be assumed to be liquid. Unfortunately, the reverse is not true, i.e., even if a bright band is not identified, precipitation at the surface can still be liquid. Other active techniques involve polarimetric measures (Matrosov et al., 2007; Zrnic et al., 1993). While these methods are promising for ground-based observations, Matrosov et al. (2009) point out that polarimetric observations are often not available for operational weather radars. Also spaceborne cloud and precipitation radars do not currently offer polarimetric capabilities.

Snowfall intensity

Ground-based active remote sensing techniques have been widely used for many years to measure snowfall intensity. First measurements of snow size distributions together with relationships between radar reflectivity (Z) and

snowfall rate (S), so-called Z–S relations, were first reported by Gunn and Marshall (1958). In 1970, a widely used size distribution and related Z–S relation was published by Sekhon and Srivasta (1970). Since then, much progress has been made in understanding the relation between radar backscatter and snowfall, as well as its inherent uncertainties. The variability in Z–S relations is largely driven by ice particle habit (shape) and size distribution – both highly variable in space and time. Various recent publications have theoretically assessed single particle and size distribution, averaged scattering properties, and their variability (Hong, 2007; Kim, 2006; Liu, 2004; Matrosov, 2007; Petty and Huang, 2010).

With the advent of CloudSat (Stephens et al., 2008) and its 94 GHz Cloud Profiling Radar (CPR), first global remote sensing observations of snowfall have become possible. Based on the theoretical work referenced above, Liu (2008), Kulie and Bennartz (2009), and Hiley et al. (2010) have studied snowfall occurrence and intensity using CloudSat's CPR. Exemplarily, zonally averaged results are shown in the lower panel of Figure 1. The uncertainty range given is solely due to uncertainties in choice of ice scattering models.

Spaceborne passive remote sensing of snow intensity or snowfall rate is hampered not only by the aforementioned uncertainties in ice scattering but also by relatively broad weighting functions and the strong and highly variable contribution of surface emission over heterogeneous, potentially snow- or ice-covered surfaces. Despite these issues in the last few years, some progress has been made regarding passive microwave estimates of snowfall at the ground (Kim et al., 2008; Noh et al., 2006, 2009; Skofronick-Jackson et al., 2004). While these studies show promising results, it is unclear at the point of writing to what extent the various issues related mostly to ice particle scattering and surface emissivity can be disentangled. More studies are needed to establish bounds on the uncertainty of passive remote sensing techniques for precipitation. A potentially interesting research venue might also lie in the combination of active and passive observations.

Summary

Over the last years, our capabilities to remotely sense snowfall have significantly increased. Significant progress has been made in three areas. Firstly, new in situ and ground-based remote sensing techniques allow us to better specify snowfall size distribution, fall speed, and particle shapes. Secondly, our understanding of basic interactions between nonspherical frozen particles and the radiation field has advanced considerably. Thirdly, with the advent of spaceborne radars and improved radiometers, it has become feasible to study snowfall globally from space. While initial studies show very promising results, many of the related retrieval techniques are still in their infancy. In the future, significant effort will have to be put into characterizing retrieval accuracies and uncertainties.

Bibliography

- Bennartz, R., 2007. Passive microwave remote sensing of precipitation at high latitudes. In Levizziani, V., Levizziani, V., Turk, J., and Bauer, P. (eds.), *Measuring Precipitation from Space – EURAINSAT and the Future*. Dordrecht: Springer, pp. 165–178. 745 p. ISBN 978-1-4020-5834-9.
- Fabry, F., and Zawadzki, T., 1995. Long-term radar observations of the melting layer of precipitation and their interpretation. *Journal of the Atmospheric Sciences*, **52**, 838–851.
- Gunn, K. L. S., and Marshall, J. S., 1958. The distribution with size of aggregate snowflakes. *Journal of Meteorology*, **15**, 452–461.
- Hiley, M., Kulie, M., and Bennartz, R., 2010. Uncertainty analysis for CloudSat snowfall retrievals. *Journal of Applied Meteorology and Climatology*, **50**(2), 399–418.
- Hong, G., 2007. Radar backscattering properties of nonspherical ice crystals at 94 GHz. *Journal of Geophysical Research-Atmospheres*, **112**, D22203, doi:10.1029/2007jd008839.
- Kim, M. J., 2006. Single scattering parameters of randomly oriented snow particles at microwave frequencies. *Journal of Geophysical Research-Atmospheres*, **111**, D14201, doi:10.1029/2005jd006892.
- Kim, M. J., Weinman, J. A., Olson, W. S., Chang, D. E., Skofronick-Jackson, G., and Wang, J. R., 2008. A physical model to estimate snowfall over land using AMSU-B observations. *Journal of Geophysical Research-Atmospheres*, **113**, D09201, doi:10.1029/2007jd008589.
- Kulie, M. S., and Bennartz, R., 2009. Utilizing spaceborne radars to retrieve dry snowfall. *Journal of Applied Meteorology and Climatology*, **48**, 2564–2580, doi:10.1175/2009jamc2193.1.
- Liu, G. S., 2004. Approximation of single scattering properties of ice and snow particles for high microwave frequencies. *Journal of the Atmospheric Sciences*, **61**, 2441–2456.
- Liu, G. S., 2008. Deriving snow cloud characteristics from CloudSat observations. *Journal of Geophysical Research-Atmospheres*, **113**, D00A09, doi:10.1029/2007JD009766.
- Matrosov, S. Y., 2007. Modeling backscatter properties of snowfall at millimeter wavelengths. *Journal of the Atmospheric Sciences*, **64**, 1727–1736, doi:10.1175/jas3904.1.
- Matrosov, S. Y., Clark, K. A., and Kingsmill, D. E., 2007. A polarimetric radar approach to identify rain, melting-layer, and snow regions for applying corrections to vertical profiles of reflectivity. *Journal of Applied Meteorology and Climatology*, **46**, 154–166.
- Matrosov, S. Y., Campbell, C., Kingsmill, D., and Sukovich, E., 2009. Assessing snowfall rates from X-band radar reflectivity measurements. *Journal of Atmospheric and Oceanic Technology*, **26**, 2324–2339, doi:10.1175/2009jtecha1238.1.
- Noh, Y. J., Liu, G. S., Seo, E. K., Wang, J. R., and Aonashi, K., 2006. Development of a snowfall retrieval algorithm at high microwave frequencies. *Journal of Geophysical Research-Atmospheres*, **111**, D22216, doi:10.1029/2005jd006826.
- Noh, Y. J., Liu, G. S., Jones, A. S., and Haar, T. H. V., 2009. Toward snowfall retrieval over land by combining satellite and in situ measurements. *Journal of Geophysical Research-Atmospheres*, **114**, D24205, doi:10.1029/2009jd012307.
- Petty, G. W., and Huang, W., 2010. Microwave backscatter and extinction by soft ice spheres and complex snow aggregates. *Journal of the Atmospheric Sciences*, **67**, 769–787, doi:10.1175/2009jas3146.1.
- Sekhon, R. S., and Srivasta, R. C., 1970. Snow size spectra and radar reflectivity. *Journal of the Atmospheric Sciences*, **27**, 299–307.
- Skofronick-Jackson, G. M., Kim, M. J., Weinman, J. A., and Chang, D. E., 2004. A physical model to determine snowfall over land by microwave radiometry. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 1047–1058.

- Stephens, G. L., Vane, D. G., Tanelli, S., Im, E., Durden, S., Rokey, M., Reinke, D., Partain, P., Mace, G. G., Austin, R., L'Ecuyer, T., Haynes, J., Lebsock, M., Suzuki, K., Waliser, D., Wu, D., Kay, J., Gettelman, A., Wang, Z., and Marchand, R., 2008. CloudSat mission: performance and early science after the first year of operation. *Journal of Geophysical Research-Atmospheres*, **113**, D00A18, doi:10.1029/2008jd009982.
- Zrníc, D. S., Balakrishnan, N., Ziegler, C. L., Bringi, V. N., Aydin, K., and Matejka, T., 1993. Polarimetric signatures in the stratiform region of a mesoscale convective system. *Journal of Applied Meteorology*, **32**, 678–693.

Cross-references

[Microwave Radiometers](#)
[Radars](#)
[Radiation, Volume Scattering](#)
[Rainfall](#)
[Weather Prediction](#)

SOIL MOISTURE

Yann Kerr
 CNES/CESBIO, Toulouse, France

Synonyms

Water content; Wetness

Definition: what is soil moisture?

Soil moisture usually refers to the amount of water stored in the soil.

Any soil can absorb a given amount of water before being saturated. It is common knowledge that different types of soils will have different types of behavior. The best example is probably the difference between a sandy beach and a clay patch in their dry and wet states. This is mainly due to the size of the soil's particles and the way water can lodge itself between them. Actually, water locates itself between the particles of soil, either very close and with a strong link (bound water) or more loosely (free water). Typically, the free water will "move" more easily (percolation, lateral flow, or evaporation) in the ground than the bound water which is more difficult to extract. The two types of status for water in soils will have an influence on the behavior of soils at microwave frequencies.

So different soils will have different "water-holding capacities" and will evolve, depending on the outside environment, between dry and wet. The most basic way of qualifying soil moisture is to range either between dry and wet or with respect to its impact on vegetation. The terms used then are the wilting point (when vegetation starts to suffer, meaning water is difficult to extract: there is only bound water left) and the field capacity (saturation, i.e., when water is so abundant that the soil cannot hold any more and water either accumulates over the surface (ponding) or runs away (runoff)).

In all of the above, soil moisture refers implicitly to the near-surface soil moisture. Actually, depending on the use of such information, soil moisture may refer to different quantities.

The most usual distinction is made between surface soil moisture and root zone soil moisture.

Surface soil moisture corresponds to soil moisture in the first centimeters of the soil. It is the soil moisture that is most visible when walking around, which interacts directly with the atmosphere (evaporation) and which drives infiltration and hence runoff in case of rain events. However, most plants have roots near the surface but also in deeper layers (depending on soil depth), so vegetation growth and health is directly linked to the water available in the root zone. The root zone is very close to what is referred to in hydrology as the vadose or unsaturated zone.

Finally, there might be another layer of stored water, deeper, the saturated zone, or water table. This layer is used by the deepest roots of trees and by man-made wells.

Just to be exhaustive one must remember that, when dealing with mass (water) transfers between the atmosphere and the soil, there are other areas where water is stored and that have an influence:

- Water stored in the vegetation, which is pumped from the soil and can be evaporated into the atmosphere through respiration/transpiration
- Water stored above the surface (lakes, rivers, ponds, snow/ice), which can be evaporated (sublimated) or can percolate or even runoff
- Water intercepted by vegetation (during rain events or as dew), which may also evaporate, be absorbed by the leaves, or eventually fall to the ground

Introduction: why measure soil moisture?

Weather forecast and extreme event prediction

Water is one of the key elements that sustain life on Earth. It is used by most of the fauna and flora and most living bodies that are mainly made of water. Consequently, humankind has always been interested in water availability. For eons, life has been organized by the cycles of vegetation and the corresponding rain cycles, with their direct impact on crop yield: predicting whether the next season would be rather wet or dry has been a constant preoccupation. In parallel to the development of techniques and the progress of science, humankind has tried to improve both its measuring skills and its forecasting capacities. If the basic concern about the weather/water availability for the next season is still a daunting challenge, our knowledge of the factors influencing the mass and energy exchanges between the surface and the atmosphere has made some progress. Models now can simulate these exchanges, taking into account the forcings (wind, solar radiation, rain, etc.) and the state of the surface (soil moisture, vegetation type and state, local slope, roughness, etc.).

Thanks to these models and observations, we have now some insights into the various factors that are crucial to

improve weather forecasts and extreme event prediction. Among them, soil moisture plays an important role as:

- A water reservoir (water storage)
- The source for water which can be evaporated into the atmosphere (mass transfer)
- A tracer of water which fell onto the surface (rain)
- A factor influencing energy budget at the surface atmosphere interface (evaporation requires energy hence induces a decrease in temperature)

To summarize, knowledge of surface soil moisture is of paramount importance for weather forecast.

Extreme events can also be better predicted when accurate values of soil moisture are used in models. An added advantage of knowing soil moisture is that large rain events (for instance, storms), depending on the soil characteristics, can lead to extreme runoff, landslides, or flooding as water infiltration is also governed by the water content of the uppermost layer.

Water resources management

Another important use of soil moisture is to get access to important information on water availability. The most obvious example is to know whether a field should be irrigated or not depending on its state, the crop growth stage and its water requirements, and the forecasted weather. This is crucial in arid or semiarid areas where irrigation is very often required, but water is not necessarily available in adequate quantities.

How to measure soil moisture

Soil moisture is traditionally measured by taking a sample of soil (in a vessel of normalized size), weighing it, drying the sample in an oven to evaporate all the water, and weighing it again. The ratio of the two measurements gives the gravimetric soil moisture. A more commonly used unit is the volumetric soil moisture which corresponds to the ratio of the sampled volume versus the volume of water it contains. The volumetric soil moisture can be inferred from the gravimetric soil moisture by simply multiplying it by the bulk density of the soil. Volumetric soil moisture is favored by many as it relates directly to remote measurements (see microwave part below).

To achieve the goals mentioned in paragraph two, it is necessary to have access to soil moisture estimates. Punctually in space and time, this is relatively easy with gravimetric sampling. However, to have measurements representative of a larger area (such as a field), the procedure is already somewhat complex as it involves a dedicated sampling strategy. Moreover, as these measurements are time consuming, regional and, even more so, global coverage are out of question. Provided one uses automatic probes (resistive, capacitive, time domain reflectometry, etc.), it is possible to achieve larger coverage and continuous measurements, but these approaches can only be confined to well-equipped and

manned sites, as they require care and maintenance. Finally, these systems carry their own problems and inaccuracies. So global monitoring of soil moisture can only rely on remote sensing from space approaches.

From space, we have access to a global approach. (All the points are measured with the same set of tool and technique, independently of countries, of ease of access, of units, or even of locally available technology.) More to the point, the measurements are by nature area integrated and thus more representative, while ground measurements are by essence very local (gravimetric samples taken a few meters apart may lead to different measurements). Conversely, if ground measurements can be very direct and accurate, measurements from space are bound to be indirect and therefore imply caveats.

Remotely sensed soil moisture, the main approaches

A large number of remote sensing approaches have been tested. For surface soil moisture, the first ones were based on shortwave measurements and on the fact that soils get darker when wet. Obviously, due to atmospheric effects and potential cloud cover, as well as vegetation cover masking effects, this approach is bound to fail in most cases. A more promising feature is linked to latent heat effects. Wet soils have a higher thermal inertia and are "cooler" than dry soils. These properties led to various trials (thermal inertia monitoring, rate of heating in the morning, surface temperature amplitude, etc.) to assess soil moisture indirectly. All those approaches proved to be somewhat disappointing due to factors inherent to optical remote sensing (atmospheric effects, cloud masking, and vegetation cover opacity) as well as to the fact that thermal infrared: (1) probes the very skin of soils and (2) the layer probed in thermal infrared is dominated by the exchanges with the atmosphere. Consequently, to infer soil moisture from such measurements, one needs to know exactly the forcings (wind, for instance, will change drastically the apparent temperature of a wet soil).

As microwave systems measure the dielectric constant of soils, which is directly related to the water content, research has quickly focused its efforts on assessing soil moisture with radars, scatterometers, or radiometers. These systems offer, when operated at low frequency, the added advantage of being all weather (measurements are not much affected by the atmosphere and clouds) and able to penetrate vegetation. They can operate at night.

Finally, in an attempt to be exhaustive, a new approach relies on measurements of the gravity field from space. As the gravity is linked to the mass, one may consider that changes in mass are mainly linked to changes in the total amount of water. In total, one should understand here: water table; water in soil layers, possibly lakes, rivers, snow, and ice; and water in vegetation and in the atmosphere. Gravimetry should thus indicate changes in the total column of water with a spatial resolution of 500 km or more. The first results certainly show a signal but its relationship with water storage has yet to be

validated and explicated. The main problem with such measurements is that they require a very large number of corrections which can be very sophisticated (for instance, taking into account the influence of tides or of the postglacial rebound). These corrections being prone to degrade the error budget in a case where the errors and corrections are of the same magnitude as the signal to be measured.

Microwaves as a tool for soil moisture monitoring: current status

The most popular approach relies on the use of synthetic aperture radars (SAR). These systems, in use since 1977 with SEASAT, offer all weather measurements with a fine spatial resolution (tens of meters). Operationally, they however suffer – as most high-resolution systems – from a rather low temporal sampling (for instance, 35 days for the European Remote Sensing (ERS) satellite) which is not really compatible with hydrologic requirements or weather forecast models. But the most adverse characteristic of SAR is the coherent nature of the signal itself and the interactions with the scattering medium. SAR images are affected by speckle and by the scattering at the surface. The scattering can be due to the vegetation cover (distribution of water in the canopy) or the soil's surface (surface scattering when wet and volume scattering). The direct consequence of these perturbations is a signal at least as sensitive to surface roughness as to moisture itself (see also Wigneron et al., 1999), not mentioning vegetation. Obviously, these effects are frequency dependent. All these inherent difficulties might explain why no real soil absolute moisture mapping was done by the several SAR that have flown since 1977. To avoid the roughness and vegetation perturbations, an approach relying on change detection, hence relative, has been used with relative success (Moran et al., 2002). However, temporal coverage is still often an issue. The use of scatterometers offers an interesting trade-off. The spatial resolution is much coarser (tens of kilometers) but with a much wider swath allowing reasonably frequent coverage (every 4–6 days on average). It also offers the added advantage of being less subject to speckle (averaging). Consequently, several authors routinely produce soil wetness maps of many areas of the world with scatterometers. The effect of vegetation is however still significant and actually corresponds to most of the signal, as the currently available frequencies are C-band on (ERS-1) and higher. So the most interesting results were obtained over arid and semiarid regions, where vegetation and soil moisture are very highly correlated anyway. The influence of surface roughness is also significant and it is best dealt with by using change detection.

The last possibility in the microwave domain is to use radiometers. The technique is old and well mastered as many sensors, notably sounders, rely on passive microwaves. To infer soil moisture, these systems are bound to offer the best compromise if used at low frequency, as

demonstrated in the early 1970s with the very short SKYLAB mission. However, to be efficient, one needs to work in a protected frequency band to avoid unwanted man-made emissions and radio frequency interferences (RFIs) and to be sensitive to soil moisture while atmosphere is transparent and vegetation plays a limited role. At L-band, the emissivity may vary from almost 0.5 for a very wet soil to almost 1 for a very dry one, giving a range of 80–100 K for an instrument sensitivity usually of the order of 1 K. As the signal is not coherent, surface roughness and vegetation structure play a reduced role when compared to active systems. So one may wonder why L-band radiometry was not used extensively before when it has been proved to be most efficient during ground and airborne measurements (Schmugge et al., 1988). This is due to an inherent limitation: The spatial resolution is proportional to the antenna diameter and inversely proportional to the wavelength. At 21 cm, to achieve a 40 km resolution from an altitude of 750 km requires an antenna of about 8 m in diameter which is a very significant technical challenge. So the research was performed with higher frequency systems as available on the scanning multichannel microwave radiometer, SMMR (6.6 GHz) (Kerr and Njoku, 1990); the special sensor microwave imager, SSM/I (19 GHz); and now the advanced microwave scanning radiometer, AMSR-E (6.8 GHz) (Njoku and Li, 1999). Even though the frequency is not adapted, good results were obtained with SMMR (in spite of a very poor resolution due to important side lobes) and now AMSR-E (Njoku and Li, 1999). The limitations are mainly linked to the fact that the vegetation becomes rapidly opaque, the frequency is not protected and thus bound to be polluted by RFIs, and the single-angular measurement makes it difficult – in several cases – to separate vegetation and soil contributions to the signal.

The step forward

Considering the necessity to make L-band measurements, several approaches have been tested to overcome the antenna-size issue. The first was initiated in the early 1990s with the idea to apply radio astronomy techniques (very large arrays) and very large baseline interferometers to Earth remote sensing. The one-dimensional concept, electronically scanned thinned array radiometer (ESTAR) was implemented as an aircraft version and proved to fulfill the requirements (Le Vine et al., 1994). It is a system, deployable in space as a sort of large rake, that offers – at the cost of a reduced sensitivity – an acceptable spatial resolution. In parallel, another approach, using inflatable (or umbrellalike) deploying technology, was studied at the Jet Propulsion Laboratory (JPL). Both concepts were proposed without success to space agencies on several occasions. The concepts appeared to be complex to deploy and to run, or to offer too limited measurements (single angle and frequency). By 1991, a small group started to work for ESA on the development of a similar instrument working in two dimensions (Goutoule et al., 1996).

The concept was named Microwave Imaging Radiometer with Aperture Synthesis (MIRAS) and an airborne prototype was made and operated (Bayle et al., 2002). From then on, the concept evolved into a more tailored instrument which was proposed to the European Space Agency (ESA) (Kerr, 1998) in the framework of the Earth Explorer Opportunity mission under the name of Soil Moisture and Ocean Salinity (SMOS) mission. The mission was selected and is now underway. It is an ESA-led project with contributions from France and Spain. SMOS is scheduled for a launch in 2009 (Kerr et al., 2001). Similarly, a mission proposal was submitted to NASA, the Hydrospheric State Mission (Hydros) (Entekhabi et al., 2004). This mission relied on a deployable and rotating antenna linked to both a radiometer and a scatterometer. It is currently being investigated under the name of Soil Moisture Active and Passive (SMAP).

SMOS is a Y-shaped instrument consisting of 69 elementary antennas regularly spaced along the arms providing at each integration step a full image (about $1,000 \times 1,200$ km) at either two polarizations or full polarization of the Earth's surface (Kerr et al., 2001). The average ground resolution is 43 km and the globe is fully imaged twice (ascending and descending orbits) every 3 days at 6 am and 6 pm local solar time (equator crossing time). As the satellite travels on the orbit, any point of the surface is imaged at several angles, giving the angular signature of the pixel. The beauty of the concept is thus that a reasonable spatial resolution is obtained at the cost of a reduced sensitivity. Moreover, the pixels are viewed frequently at different angles and polarizations. The angular information is then used to separate the different contributions (soil – vegetation) to the signal (Wigneron et al., 2000).

Root zone soil moisture

The big caveat of the remotely sensed soil moisture is that the direct measurement only concerns the surface layer. For instance, at X-band, a few millimeters and, at L-band, 4–5 cm are probed on average (depending on soil characteristics and condition). It is necessary however, for several applications, to know the available water in the unsaturated zone. Only one direct approach can currently be considered: using even lower frequencies (wavelengths of several meters) so as to reach deeper layers. This leads to large problems in terms of spatial resolution (a few hundred kilometers) as well as problems linked to ionospheric effects. So this option is not feasible now. The indirect approaches could be either to use gravity change approaches (if really validated and provided very coarse resolutions – hundreds of kilometers – are acceptable) or to rely on assimilation techniques, i.e., to use models to infer – from regular surface measurements and forcing conditions – what the root zone soil moisture is. The approach has been validated both using simulations and using ground data. The real limitations are linked to the models' performances and to the input data quality.

Expression of needs

Obviously, apart from the fact that the unsaturated zone is only partly probed, there are some requirements that are not fully fulfilled. The main one is the spatial resolution. Some needs, notably in hydrology, can only be resolved by having a better spatial resolution while still retaining a high temporal sampling. From space, this is not straightforward, but the most promising solution is probably to use external information to redistribute the area's average moisture within the pixel: the so-called disaggregation. Several studies recently proved the feasibility of the approach with SMOS data (Pellenq et al., 2003; Merlin et al., 2006) and now the real-life validation has only to be performed.

Caveats

This is not to say all problems have been solved. There are still a number of outstanding issues which will require attention before an accurate and global soil moisture product is routinely delivered. Some problems, such as RFI, can be general issues, especially if protection is reduced in the future, which is a concern. The specific issues identified are currently being tackled and several references in the literature identify them, but, obviously, as long as the actual data (SMOS or any other) are not available, definitive conclusions and/or solutions will not be available and some unexpected issues might arise. Currently, the following issues are well identified. The most stringent is the pixel heterogeneity with components which may have very significant differences in behavior. The presence of free water within the pixel, for instance, has to be very accurately known (better than 2 %) to reach the overall accuracy of 4 % vol. in soil moisture. And water bodies can be variable as a function of season or weather. Vegetation is not totally opaque at L-band, and when the integrated water content is above $4\text{--}5 \text{ kg/m}^2$, soil moisture retrievals will be difficult and approximate. Therefore, forested canopies will impact the signal. It may be noted at this level that recent studies showed that the main contributor at L-band for forest was the branch which does not evolve so rapidly (Ferrazzoli and Guerriero, 1996). Litter on the ground may behave as blackbody, masking strongly the soil's signal. During rain events, water interception by the canopy might artificially increase the apparent vegetation's water content. Topography will induce an altered angular behavior; snow and frozen soils will induce different signals which, if not accounted for, will produce wrong estimates. Urban areas and rocks are not fully assessed in terms of emissivity. Finally, and generally speaking, good retrieval will require some a priori knowledge of the surface cover, and state and the quality of the retrievals will be linked to the quality of the input data. It must also be noted that systems like SMOS will bring inherent specificities and complexities such as image reconstruction which is still a challenging point.

Conclusion and perspectives

Soil moisture has been for a long time a very specific variable. Though well identified, there were so few global measurements that models do not use it per se. However, after many unsuccessful attempts, a real soil moisture mission is now underway. It should enable the community to have, finally, access to global fields of soil moisture. Until launched and commissioned, the concept still has to be proved but the elements available can make one very confident to such an extent that an operational SMOS follow-on is currently being studied. Nevertheless, if SMOS answers some questions, (it) still does not fulfill all existing needs and ways forward must be sought. The most important is probably to improve the spatial resolution and there the SMOS concept is close to an optimum, as increasing the arm's length will improve the spatial resolution but degrade significantly the sensitivity, to the point where it would not be useful anymore. So other techniques, such as disaggregation, will have to be found. To be more efficient, a SMOS-like instrument might gain either from being multifrequency or from having a coupled active system. To test those options, we will be expected to test these solutions using existing sensors (Advanced SCATterometer (ASCAT) and AMSR-E) when SMOS is operating. Another approach to improve spatial resolution could be to use even larger antennas, depending on the possibility of deploying them efficiently in space. In that case, to resolve the ambiguities, it will probably be necessary to improve the system by adding other frequencies while keeping the active source. This might lead to addressing the cryosphere as well, another key element in the global water and energy budget of the planet.

It was stated in the SMOS proposal that the concept, though challenging, would open a new field with new measurements – soil moisture – made with a new type of sensors, paving the way for operational monitoring of water in soils. With the inception of the SMOS mission, this step is taken, opening a whole avenue of scientific challenges and making the long-awaited tool for water resources and water cycle monitoring a closer possibility.

Bibliography

- Bayle, F., Wigneron, J.-P., Kerr, Y. H., Waldteufel, P., Anterrieu, E., Orlhac, J.-C., Chanzy, A., Marloie, O., Bernardini, M., Sobjaerg, S., Calvet, J.-C., Goutoule, J.-M., and Skou, N., 2002. Two-dimensional synthetic aperture images over a land surface scene. *IEEE Transactions on Geoscience and Remote Sensing*, **40**(3), 710–714.
- Entekhabi, D., Njoku, E. G., Houser, P., Spencer, M., Doiron, T., Kim, Y., Smith, J., Girard, R., Belair, S., Crow, W., Jackson, T. J., Kerr, Y. H., Kimball, J. S., Koster, R., McDonald, K. C., O'Neill, P. E., Pultz, T., Running, S. W., Shi, J., Wood, E., and van Zyl, J., 2004. The hydrosphere state (hydros) satellite mission: an earth system pathfinder for global mapping of soil moisture and land freeze/thaw. *IEEE Transactions on Geoscience and Remote Sensing*, **42**(10), 2184–2195.

- Ferrazzoli, P., and Guerriero, L., 1996. Passive microwave remote sensing of forests: a model investigation. *IEEE Transactions on Geoscience and Remote Sensing*, **34**, 433–443.
- Goutoule, J. M., Anterrieu, E., Kerr, Y. H., Lannes, A., and Skou, N., 1996. *MIRAS Microwave Radiometry Critical Technical Development*. Toulouse: MMS.
- Kerr, Y. H., 1998. *The SMOS Mission: MIRAS on RAMSES. A Proposal to the Call for Earth Explorer Opportunity Mission*. Toulouse: CESBIO.
- Kerr, Y. H., and Njoku, E. G., 1990. A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space. *IEEE Transactions on Geoscience and Remote Sensing*, **28**(3), 384–393.
- Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J.-M., Font, J., and Berger, M., 2001. Soil moisture retrieval from space: the soil moisture and ocean salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(8), 1729–1735.
- Le Vine, D. M., Griffis, A. J., Swift, C. T., and Jackson, T. J., 1994. ESTAR: a synthetic aperture microwave radiometer for remote sensing applications. *Proceedings of the IEEE*, **82**, 1787–1801.
- Merlin, O., Chehbouni, A. G., Kerr, Y. H., and Goodrich, D., 2006. A downscaling method for distributing surface soil moisture within a microwave pixel: application to monsoon '90 data. *Remote Sensing of Environment*, **101**, 379–389.
- Moran, M. S., Hymer, D. C., Qi, J., and Kerr, Y., 2002. Comparison of ERS-2 SAR and Landsat TM imagery for monitoring agricultural crop and soil conditions. *Remote Sensing of Environment*, **79**(2–3), 243–252.
- Njoku, E. G., and Li, L., 1999. Retrieval of land surface parameters using passive microwave measurements at 6–18 GHz. *IEEE Transactions on Geoscience and Remote Sensing*, **37**(1), 79–93.
- Pellenq, J., Kalma, J., Boulet, G., Saulnier, G.-M., Wooldridge, S., Kerr, Y., and Chehbouni, A., 2003. A disaggregation scheme for soil moisture based on topography and soil depth. *Journal of Hydrology*, **276**(1–4), 112–127.
- Schmugge, T. J., Wang, J. R., and Asrar, G., 1988. Results from the push broom microwave radiometer flights over the Konza Prairie in 1985. *IEEE Transactions on Geoscience and Remote Sensing*, **26**(5), 590–597.
- Wigneron, J. P., Ferrazzoli, P., Calvet, J. C., Kerr, Y. H., and Bertuzzi, P., 1999. A parametric study on passive and active microwave observations over a soybean crop. *IEEE Transactions on Geoscience and Remote Sensing*, **37**(6), 2728–2733.
- Wigneron, J.-P., Waldteufel, P., Chanzy, A., Calvet, J. C., and Kerr, Y., 2000. Two-D microwave interferometer retrieval capabilities of over land surfaces (SMOS mission). *Remote Sensing of Environment*, **73**(3), 270–282.

Cross-references

[Agriculture and Remote Sensing](#)
[Climate Data Records](#)
[Climate Monitoring and Prediction](#)
[Data Assimilation](#)
[Global Climate Observing System](#)
[Irrigation Management](#)
[Land Surface Temperature](#)
[Land-Atmosphere Interactions, Evapotranspiration](#)
[Microwave Dielectric Properties of Materials](#)
[Microwave Radiometers](#)
[Microwave Radiometers, Conventional](#)
[Microwave Radiometers, Interferometers](#)
[Microwave Surface Scattering and Emission](#)
[Radar, Scatterometers](#)
[Radars](#)
[Radiation, Volume Scattering](#)
[Radiative Transfer, Theory](#)

Rainfall
Remote Sensing, Physics and Techniques
Soil Properties
Surface Water
Water and Energy Cycles
Water Resources

SOIL PROPERTIES

Alfredo Huete
Plant Functional Biology and Climate Change Cluster,
Faculty of Science, University of Technology, Sydney,
NSW, Australia

Synonyms

Pedology; Pedosphere

Definition

Soils. Soils are three-dimensional living bodies, with spatially variable biologic, physical, and chemical properties, that form the outer skin of the Earth's terrestrial surface.

Soil formation. Soils form slowly over time and develop distinguishing properties as a function of climate, geologic and organic parent materials, topography, time, vegetation type, and land use history.

Soil profile. The vertical depth of a soil body varies from a few centimeters up to several meters and contains a series of soil horizons. The surface layers are termed the "O" (organic) or "A" (mineral) horizons, while a lower zone of clay accumulation is the "B" horizon, and the lowest zone that interfaces with the parent material is the "C" horizon.

Introduction

The thin biologically active soil layer of the Earth's terrestrial surface is an integral part of the biosphere, regulating biogeochemical and hydrologic cycling of matter and energy and affecting land productivity and biodiversity. Soils perform numerous ecologic functions, as they buffer and filter pollutants that impact water quality, serve as source and sink of greenhouse gases, and partition rainfall between infiltration and runoff. Soils are home to the richest biodiversity on the planet that includes countless microorganisms that perform biochemical transformations, vital to ecosystem functioning, such as nitrogen fixation and organic matter decomposition. Reliable information on the spatial distribution of soil properties along with a better understanding of fundamental soil processes are vital to achieve sustainable land management of ecosystem services and to solve many environmental challenges facing society.

Soils exhibit great spatial variability as a result of the interactions of climate, topography, parent material, and organisms acting on the soil body over time (Amundson and Jenny, 1997). Human activities and land use history

also play important roles in determining soil properties and additionally contribute to the heterogeneous soil spatial patterns. Remote-sensing systems, with their synoptic, multi-scale, and repetitive coverage of the land surface, provide an ideal means of mapping the spatial variation and biogeophysical properties of soils for environmental and natural resource management purposes.

Soil optical properties

Soil optical properties have been extensively studied in laboratory and field measurements and found to be primarily dependent on soil biogeochemical composition, soil surface geometric-optical scattering, and surface moisture status, as reviewed in Baumgardner et al. (1985). Generally, the spectral composition and the strength of the optical signal measured are determined by the minerals and organic compounds that coat the surface layer of soil particles (Stoner and Baumgardner, 1981). In contrast to the strong absorption spectra of pure mineral samples, soil absorption features are only broadly defined as soils consist of aggregate mixtures of many mineral and organic constituents. Many quantitative tools for extracting information about the physical and biochemical characteristics of soils have been developed and recognized as the field of soil spectroscopy (Irons et al., 1989; Ben-Dor et al., 1999). As an example, Ben-Dor and Bannin (1994) utilized a visible and near-infrared analysis scheme to predict a wide variety of soil chemical constituents, including CaCO_3 , Fe_2O_3 , Al_2O_3 , SiO_2 , free iron oxides, and K_2O , from high-resolution spectra of arid and semiarid soils.

In field radiometric measurements of in situ soils, there are strong optical-geometric interactions resulting from the angles in which the sun illuminates, and the sensor views, the surface. In addition, the atmosphere influences the relative proportions of direct (sun) and diffuse (sky) illumination on the soil surface. Particle size distribution and surface-height variation (roughness) are some of the most important factors influencing the reflectance properties of in situ soils. The bidirectional reflectance distribution function (BRDF) describes the manner in which surfaces scatter radiation across all sun-surface-sensor view geometries and can be used, through models, to derive geometric descriptors of a soil, such as size, shape, and orientation of surface "roughness" elements (Irons et al., 1992). There is a general decrease in reflectance with increasing surface roughness as coarse aggregates contain a lot of inter-aggregate spaces and "light traps." With the shortest wavelengths most affected, optical-geometric interactions alter a soil's spectral signature and the inferences made of soil biogeochemical properties such as soil color and mineralogy.

When integrated across a hemisphere, the BRDF provides "albedo," the ratio of shortwave (0.4–4 μm) radiant energy scattered in all directions to the incident irradiance on the surface. Albedo is a fundamental variable in energy balance studies, climate modeling

studies, and soil degradation studies, e.g., smooth, crusted, and structureless soils generally reflect more energy and are brighter than well-structured soils. Soil moisture also has a strong influence on the amount and composition of reflected energy from a soil with a general decrease in reflectance, proportional to the thickness and energy status of the adsorbed water. Soil moisture and soil roughness impart strong, or bright, microwave signals, and image-differencing techniques can be used to separate the dynamic soil moisture signal from the relatively static soil roughness signal (Sano et al., 1998). More recently developed **LiDAR** (Light Direction and Ranging) technology enables land surface roughness to be measured, on the order of centimeters, and offers promise in mapping terrain stability and hazard analysis, such as landslides (Jones, 2006).

Space- and airborne imagery

Remotely sensed imagery from air- and spaceborne sensors integrates much of the knowledge learned from laboratory and field spectroscopy into the spatial domain and to the much more heterogeneous landscape scale. There is a wide array of airborne and satellite sensor systems currently available to study soils at the landscape level, encompassing a wide range of spatial and temporal resolutions, number of spectral bands and bandwidths, sun-view geometries, and polarization states. The derivation of soil information within landscapes may involve (1) direct measurements of exposed soils; (2) extraction of soil information from partially vegetated, open canopies; and (3) inferring soil properties from measurements of the vegetation layer. In all cases, correction and filtering of the image data are required to minimize atmosphere contamination, clouds, topography effects, BRDF influences, and sensor noise characteristics.

Galvão et al. (2008) demonstrated how a subset of exposed soil pixels from airborne AVIRIS and spaceborne ASTER imagery, collected in Campo Verde, Brazil, could be used to study several soil physical and chemical weathering properties and their relationship to landscape position. Spectral soil indices, often in the form of band ratios, and spectral mixture analysis (SMA) have been used to ascertain soil information where vegetation and litter residues mask portions of the soil surface and pure soil pixels are not observable. SMA unmixes pixels into their respective soil, vegetation, and litter signal contributions, useful in soil and vegetation mapping, land degradation, soil erosion, and land cover conversion studies (Adams et al., 1986; Smith et al., 1990; Ustin et al., 1993). As an example, Palacios-Orueta et al. (1999) used a sequential spectral mixture analysis technique that utilized laboratory-derived “training” vectors to extract soil organic matter and iron sources of spectral variation. Alternatively, one may also infer soil type and properties from remote-sensing measures of foliar chemistry, a technique known as *geobotany* (see Ustin et al., 1999).

Landscape soil processes

There have been considerable accomplishments made in understanding dynamic soil processes at the landscape scale with remote-sensing data, particularly in the areas of soil degradation, soil erosion, salinization, soil contamination, drought, carbon cycling, and climate and land use impacts. Soils are vulnerable and susceptible to poor land use practices that can yield soils with different physical properties and characteristics and thereby impact on their ability to cycle water and nutrients for sustained plant growth.

Soil degradation

Soil degradation is a major environmental concern that impacts on critical environmental issues such as food security, water quality, loss of biodiversity, and global climate change. Degraded soils exhibit an appreciable loss of soil organic matter with reduced fertility and plant productivity, deteriorated soil structures, and increases in albedo, thus altering carbon cycling and the land surface energy balance. Remote sensing provides repeatable and verifiable techniques for monitoring and assessing the spatial extent and severity of soil degradation. This is accomplished through satellite-based indicators of soil degradation that measure the loss of vegetative cover, increases in albedo, reduced soil organic matter levels, salinization, wind and water erosion, and soil crusting. Satellite data are increasingly being utilized to monitor the albedo of arid and semiarid lands given the importance of albedo as an indicator of land degradation and as a physical parameter with possible impacts on climate. Spatial heterogeneity indices from satellite images are also sensitive indicators of landscape instability, with increases in variance due to soil degradation and erosion (Pickup and Nelson, 1984).

An important threat posed by soil degradation is soil erosion, in which the upper and most fertile soil layers are removed by wind and water forces and deposited into waterways and coastal zones, resulting in loss of arable land and the silting of reservoirs. Windblown atmospheric dust may also pose public health concerns. Soil erosion is evident in remote-sensing imagery when subsoil characteristics are exposed, which contrast with the “intact” surface soil spectral signals. In very fine spatial resolution imagery, such as from QuickBird, detailed information on linear erosion features such as gullies, rills, and sand dune formations can be directly observed.

Soil carbon

There is great interest in analyzing the importance of soils as a source or sink of greenhouse gases such as carbon dioxide, methane, and nitrous oxide. The soil carbon pool is approximately twice that in the atmosphere and three times that in vegetation (Smith et al., 2008). Although the response of soils to global warming is critical, reliable estimates of carbon reserves in undisturbed soils are not available, and data on rates of carbon loss following

disturbances are lacking (Lal, 1997). Soils subject to accelerated decomposition of organic matter tend to release carbon dioxide and thus contribute to the enhanced greenhouse effect, while areas being revegetated and enriched with organic matter can absorb and sequester quantities of carbon that are extracted from the atmosphere through photosynthesis. In the context of global sustainability, it is essential to understand how the source/sink function of soils can be managed and controlled to mitigate the impact of climate change (Post et al., 1982).

Soils also play important roles in water quality, sediment and pollutant discharges into water systems, and environmental remediation efforts. The soil has the ability to filter and break down toxic waste materials and many industrial by-products and render them less harmful. New capabilities are also becoming available for soil process studies enabling the integration of environmental process models, climate data, human interactions, and Geographic Information Systems (GIS) with remote-sensing data.

Conclusion

In summary, soils exhibit tremendous spatial and dynamic variations across a landscape due to variations in moisture content, biogeochemical composition, surface roughness elements, and ongoing processes of land use, soil degradation, erosion, contamination, and global change. Although not fully developed, remote sensing remains the only viable technique to map, monitor, and manage the fragile soil resource. With over 90 % of the Earth's terrestrial surface classified as "open canopies," there is an appreciable soil component that can potentially be remotely sensed.

Many of the remote-sensing techniques used to derive soil information, however, remain undeveloped, and there are no currently available operational satellite data products on *soil properties*. In part, the extraction of soil information from landscape-scale imagery is too complex, hindering many potential end uses. As noted by Ben-Dor et al. (2008), the adoption of imaging spectroscopy techniques, although cost-effective, remains a new frontier in soil science with advances limited by the lack of availability of operational hyperspectral sensors. Nevertheless, satellite data products, such as albedo, BRDF, and vegetation indices, measure important landscape properties that are useful in deriving soil information.

Remote sensing is also only sensitive to the immediate surface, which prevents the assessments of important subsoil and root-zone soil properties. However, the soil surface is the most dynamic, biologic, and hydrologic interface that responds immediately to climatic changes and human forcings. Remote sensing can provide quantitative measurements of such changes at the surface, which are needed to address a wide range of environmental and global change issues. Further advancements are needed to infer soil properties with depth through pedotransfer functions that enable the derivation of subsoil biogeophysical products from remotely sensed surface properties.

Bibliography

- Adams, J. B., Smith, M. O., and Johnson, P., 1986. Spectral mixture modeling, a new analysis of rock and soil types at the Viking Lander 1 site. *Journal of Geophysical Research*, **91**(B8), 8098–8112.
- Amundson, R., and Jenny, H., 1997. On a state factor model of ecosystems. *BioScience*, **47**, 536–543.
- Baumgardner, M. F., Silva, L. F., Biehl, L. L., and Stoner, E. R., 1985. Reflectance properties of soils. *Advances in Agronomy*, **38**, 1–44.
- Ben-Dor, E., and Bannin, A., 1994. Visible and near infrared (0.4–1.1 μm) analysis of arid and semi arid soils. *Remote Sensing of Environment*, **48**, 261–274.
- Ben-Dor, E., Irons, J. R., and Epema, G., 1999. Soil reflectance. In Rencz, A. N. (ed.), *Remote Sensing for the Earth Sciences, Manual of Remote Sensing*, 3rd edn. New York: Wiley, pp. 111–188.
- Ben-Dor, E., Taylor, R. G., Hill, J., Demattê, J. A. M., Whiting, M. L., Chabrilat, S., and Sommer, S., 2008. Imaging spectrometry for soil applications. *Advances in Agronomy*, **97**, 321–392.
- Galvão, L. S., Formaggio, A. R., Couto, E. G., and Roberts, D. A., 2008. Relationships between the mineralogical and chemical composition of tropical soils and topography from hyperspectral remote sensing data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **63**, 259–271.
- Irons, J. R., Weismiller, R. A., and Petersen, G. W., 1989. Soil reflectance. In Asrar, G. (ed.), *Theory and Application of Optical Remote Sensing*. New York: Wiley, pp. 66–106.
- Irons, J. R., Campbell, G., Norman, J. M., Graham, D. W., and Kovalick, W. M., 1992. Prediction and measurement of soil bidirectional reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-30**, 249–260.
- Jones, L., 2006. Monitoring landslides in hazardous terrain using terrestrial LiDAR: an example from Montserrat. *Quarterly Journal of Engineering Geology & Hydrogeology*, **39**, 371–373.
- Lal, R., 1997. Soil processes and the greenhouse effect. In Lal, R., Blum, W. H., Valentin, C., and Stewart, B. A. (eds.), *Methods for Assessment of Soil Degradation*. Boca Raton, FL: CRC Press. *Advances in Soil Science*, pp. 199–212.
- Palacios-Orueta, A., Pinzón, J. E., Ustin, S. L., and Roberts, D. A., 1999. Remote sensing of soils in the Santa Monica Mountains: II. Hierarchical foreground and background analysis. *Remote Sensing of Environment*, **68**, 138–151.
- Pickup, G., and Nelson, J., 1984. Use of Landsat radiance parameters to distinguish soil erosion, stability and deposition in arid central Australia. *Remote Sensing of Environment*, **16**, 195–204.
- Post, W. M., Emanuel, W. R., Zinke, P. J., and Stangenberger, A. G., 1982. Soil carbon pools and world life zones. *Nature*, **298**, 156–159.
- Sano, E. E., Huete, A. R., Troufleau, D., Moran, M. S., and Vidal, A., 1998. Relation between ERS-1 synthetic aperture radar data and measurements of surface roughness and moisture content of rocky soils in a semiarid rangeland. *Water Resources Research*, **34**, 1491–1498.
- Smith, M. O., Ustin, S. L., Adams, J. B., and Gillespie, A. R., 1990. Vegetation in deserts: I. A regional measure of abundance from multispectral images. *Remote Sensing of Environment*, **31**, 1–26.
- Smith, P., Fang, C., Dawson, J. J. C., Moncrieff, J. B., and Smith, P., 2008. Impact of global warming on soil organic carbon. *Advances in Agronomy*, **97**, 1–43.
- Stoner, E. R., and Baumgardner, M. F., 1981. Characteristic variations in reflectance of surface soils. *Soil Science Society of American Journal*, **45**, 1161–1165.
- Ustin, S. L., Smith, M. O., and Adams, J. B., 1993. Remote sensing of ecological processes: a strategy for developing and testing ecological models using spectral mixture analysis.

In Ehleringer, J. R., and Field, C. B. (eds.), *Scaling Physiological Processes: Leaf to Globe*. San Diego: Academic, pp. 339–357.

Ustin, S. L., Smith, M. O., Jacquemoud, S., Verstraete, M. M., and Govaerts, Y. M., 1999. Geobotany: vegetation mapping in earth sciences. In Rencz, A. N. (ed.), *Remote Sensing for the Earth Sciences, Manual of Remote Sensing*, 3rd edn. New York: Wiley, pp. 189–248.

Cross-references

Land Surface Roughness
 Land-Atmosphere Interactions, Evapotranspiration
 Lidar Systems
 Soil Moisture
 Water and Energy Cycles
 Water Resources

SOLID EARTH MASS TRANSPORT

Erik Ivins
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Synonyms

Time-variable gravity of the solid Earth

Definition

Time-variable gravity of the solid Earth. Any measurement of gravity at, or below, the Earth's surface or from above by airborne or space instruments, wherein the physical origins of the variations are solely, or interactively, caused by movements, deformations, or changes in density of the solid Earth. Generally, solid Earth time variability includes, but is not limited to, fracture or poroelastic-related crustal volatile and hydrological material transport, changes of state and solid Earth's response to periodic forcing by ocean and lunisolar tides, earthquake-related stress reorganization, or by nutation and wobble of the Earth's rotation axis. Large-scale and long-term mass variability is associated with the solid creep of the mantle in response to the surface environmental load changes associated with the last global ice age (100,000–10,000 years ago). Smaller-scale mass variability driven by flow of the upper mantle is also associated with the more recent Little Ice Age surface loading, involving time scales of 50–500 years and surface vertical motions of 1–40 mm/year. Only recently have the last two phenomena become amenable to accurate measurement by modern space gravimetry.

Introduction

The study of the Earth's gravity, topography, and seismological structure has provided several of the most basic observational cornerstones that support the dynamic theories for the evolution of Earth's mountain belts, sedimentary basins, massive continental eruptive and breakup events, ocean ridges, plate motions, slab subduction, and all phenomenon that are generally related to mantle

convection (e.g., Richards and Hager, 1984; Forte and Woodward, 1997). While inferences concerning gravity change and mass transport may be made for mantle convection, the minimum relevant time scales, τ , are defined by those for plate tectonic reorganization ($\tau \geq 5$ million years). The class of solid Earth time-variable gravity that is measurable by satellite remote-sensing methods is shorter by at least a factor of 10^4 ($\tau \leq 50$ years). Experts in near-Earth orbital dynamics developed early measurement concepts focused on simple drifts and periodic patterns in well-tracked orbital elements (Paddack, 1967; Kozai, 1968). The launch of two passive laser-ranging satellites, Starlette and LAGEOS-1 in 1976, marked the beginning of a new era in global gravity remote sensing. The key orbital elements proved to be sensitive to changes in planetary oblateness, and these were best explained by the presence of a slow viscoelastic flow of the mantle in response to the two-way transfer of ocean-to-continent-to-ocean ice sheet expansion and contraction that began about 100,000 years before present (year BP) and ended 7,000 year BP. The response phenomenon is known as glacial isostatic adjustment (GIA) of the combined solid and liquid Earth. GIA alone gave a reasonable explanation for the amplitude and sense of the nontidal part of variations in the Earth's gravity field (Yoder et al., 1982; Peltier, 1982; Rubincam, 1984). An excellent survey of the methods and theoretical representation of the static gravitational field prior to time-variable mapping has been given by Phillips and Lambeck (1980). These methods, including coefficient-normalization conventions, are applicable to time-varying gravity.

Passive satellites in constellation

With the passing of two decades since the launch of Starlette and LAGEOS-1, and the accumulation of tracking data for a virtual constellation of more than ten subsequent satellites, it has been possible to refine the best-fitting GIA models and to incorporate realistic scenarios for the mass balance of mountain glaciers, the Greenland and Antarctic ice sheets. A series of papers, each analyzing slightly different tracking-derived solutions for the zonal harmonic drift rates (e.g., Cheng et al., 1997), identified a statistically significant improvement in the residual fit to all zonal harmonic rates of drift by accounting for Antarctic and Greenland ice mass change. The solved-for drift rates are especially sensitive to the latitudinal dependencies of the low-degree zonal harmonics, hence natural antecedents to the study of both GIA and present-day mass balance of the continental cryosphere. James and Ivins (1997), Johnston and Lambeck (1999), Cox et al. (2001), and Tosi et al. (2005) used zonal harmonic rates of order $n \leq 9$ and showed that Antarctic mass loss at rates of -70 to 340 Gt/year ($1 \text{ Gt} = 10^9 \text{ t}$) was favored across a broad spectrum of mantle viscosity profiles. A critical aspect of the finely tuned relationships between nontidal satellite orbital element drifts to solid Earth GIA and cryospheric imbalance is that modeled

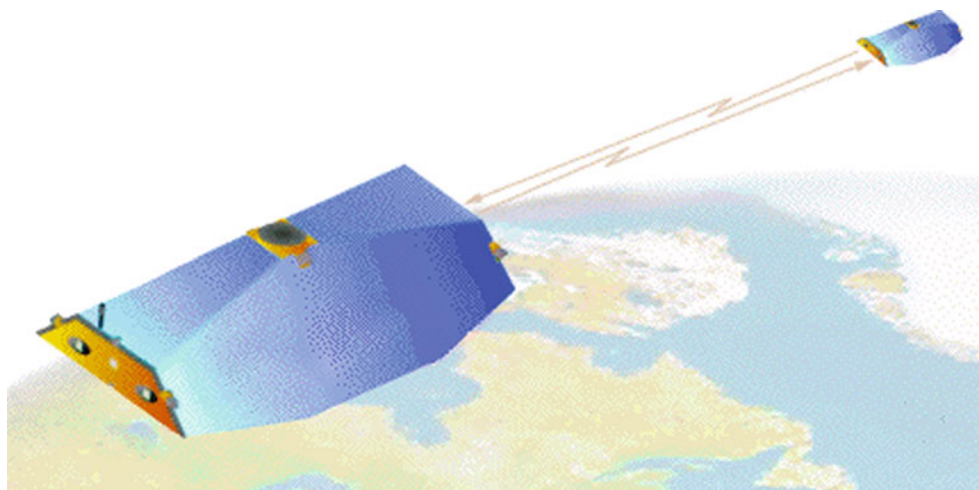
accelerations due to drag have appropriate error estimates and that the mass changes in the oceans have no large-amplitude, low-order, and long-term secular-like components (cf. Hughes and Stepanov, 2004).

Satellite-to-satellite tracking: a new era for time-variable gravity

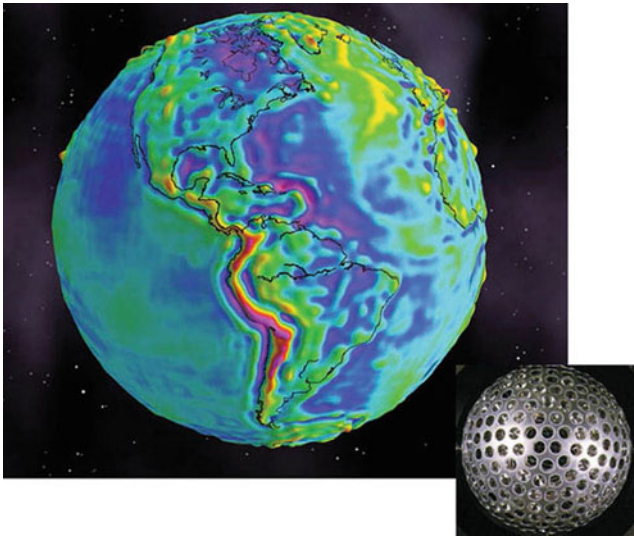
Plans for a high-resolution global gravity-mapping mission with satellites at mean orbital altitudes below 800 km were well underway by the late 1970s (e.g., Kaula, 1983). The lower orbits of nonpassive satellites required new navigational and tracking strategies. Passively laser-tracked gravity satellites in nearly circular orbit have apogees, a , ranging from 7,173 to 25,001 km and are relatively drag-free. The measurement sensitivity to surface gravity falls off as $(R_e/a)^{n+1}$ where R_e is the mean radius of the solid Earth and the harmonic degree and n is related to spatial resolution as the half wavelength $\Delta\lambda \approx R_e \pi/n$. Optimistically, the limits of spatial resolution (in latitude only) of the passive satellites are $\Delta\lambda < 2,200$ km. The limitation is actually far more severe than this. At full wavelength, $2\Delta\lambda$, say, for example, the orbits of these passive satellites are responsive to mass changes at, or beneath, the combined Central Siberian Plateau and Western Siberian Plain (representing approximately one-fifth of the Eurasian land area). Yet the rate solutions for the harmonic coefficients, $\partial_t C_{nm}$, determined by passive-satellite tracking are unable to resolve the latitudinal location of the disturbing mass, for there is equal probability that the observed orbital drifts are related to mass changes in central Canada! This drawback is related to the lack of sensitivity to nonzonal harmonics (generally, nonaxially symmetric), having $m \neq 0$. This is essentially the “Achilles’ heel” of satellite laser-ranging (SLR) strategies for monitoring time-variable gravity from space.

Multiple-satellite tracking and smaller a open a new window to resolution. In fact, the German geoscience satellite launched in 2000, CHAMP, utilizing tracking by the entire constellation of high-orbit Global Positioning System (GPS) satellites, and onboard accelerometer and star-tracking data, was capable of resolving the global static field to degree $n_{\max} \approx 72$ (Reigber et al., 2003). Although sampling in polar orbit yields a degraded longitudinal resolution, effectively limiting nonzonal resolution to $m_{\max} \approx O(n_{\max})/2$, seasonal changes in mass can unambiguously detect mass changes associated, for example, to the combined Central Siberian Plateau and Western Siberian Plain. The role that a unified single tracking reference frame plays in reducing the errors in the end-product gravity mapping should not be underestimated (Rummel et al., 2005). The polar orbit also allowed, for the first time, a global gravity map to be completed that is based on a single instrument and observing scheme, as opposed to the prior generation of merging strategies required to produce global maps. The latter always contained regionally varying quality and had no possibility of representing global time-varying properties.

The Gravity Recovery and Climate Experiment (GRACE) satellite pair determines an even higher-resolution static and time-variable field. These were launched into polar orbit in 2002 and fly in tandem, separated by about 220–240 km. The two track one another with a Ka-band radar system, in addition to using onboard accelerometers and GPS tracking (Figure 1). Although degraded in resolution by the same anisotropic sampling problem as CHAMP, the resolution recovered by a standard isotropic Gaussian filtering technique (Wahr et al., 1998) is about 400 km, and a variety of more sophisticated filtering methods may reduce this by 50–130 km, depending on the sought-after frequency component and latitude (e.g., Davis et al., 2008).



Solid Earth Mass Transport, Figure 1 GRACE satellite pair with Ka band microwave tracking links. The separation is about 220–240 km. The altitude is about 550 km and the pair has roughly 15 Earth revolutions per day. Also see http://www.nasa.gov/mov/161008main_GraceBeauty1web.mov (Graphic from http://op.gfz-potsdam.de/grace/index_GRACE.html).



Solid Earth Mass Transport, Figure 2 GRACE-based static gravity field and inset photo of the passive gravity satellite Lageos-2 equipped with retroreflectors, weighing 400 kg and having a 60 cm diameter (http://www.nasa.gov/vision/earth/lookingatearth/earth_drag.html).

The higher-accuracy and complete even-degree harmonic retrieval of the fields allowed for a highly accurate determination of modeled residual drifts in the nodes of the LAGEOS-1 and 2 satellites and the later confirmed the predicted drifts of Einstein's theory of relativity. The drifts are a result of additional space-time curvature caused by the Earth's rotation (Ciufolini and Pavlis, 2004). The passive, laser-reflecting, LAGEOS-2 satellite is shown as an inset to the GRACE-gravity anomaly map shown in Figure 2.

GRACE and CHAMP determine fields and their time variability at higher latitudes with relatively enhanced fidelity. The mapped fields are inherently less prone to errors due to the greater collection of orbital mapping data. The standard (level-2) data consist of monthly (or sub-monthly) spherical harmonic coefficients, C_{nm} , as there exist a collection of well-developed methodologies for such representation and use (e.g., Rapp and Pavlis, 1990; Petrovskaya et al., 2001; Tapley et al., 2004). Important results derived from the harmonic releases for polar latitudes include the detection of huge ice loss in the south-eastern part of Greenland (Velicogna and Wahr, 2006a) and in West Antarctica (Velicogna and Wahr, 2006b; Ramillien et al., 2006). A more sophisticated method uses the raw satellite-to-satellite range and range-rate data and accelerometer data to directly infer the location and size of perturbing masses beneath the satellites without reference to global spherical harmonics. The methods use the same basic potential field representation for lunar mascons that was developed in the late 1960s (e.g., Muller and Sjogren, 1968). These direct methods have shown great sensitivity to year-to-year ice mass loss in coastal

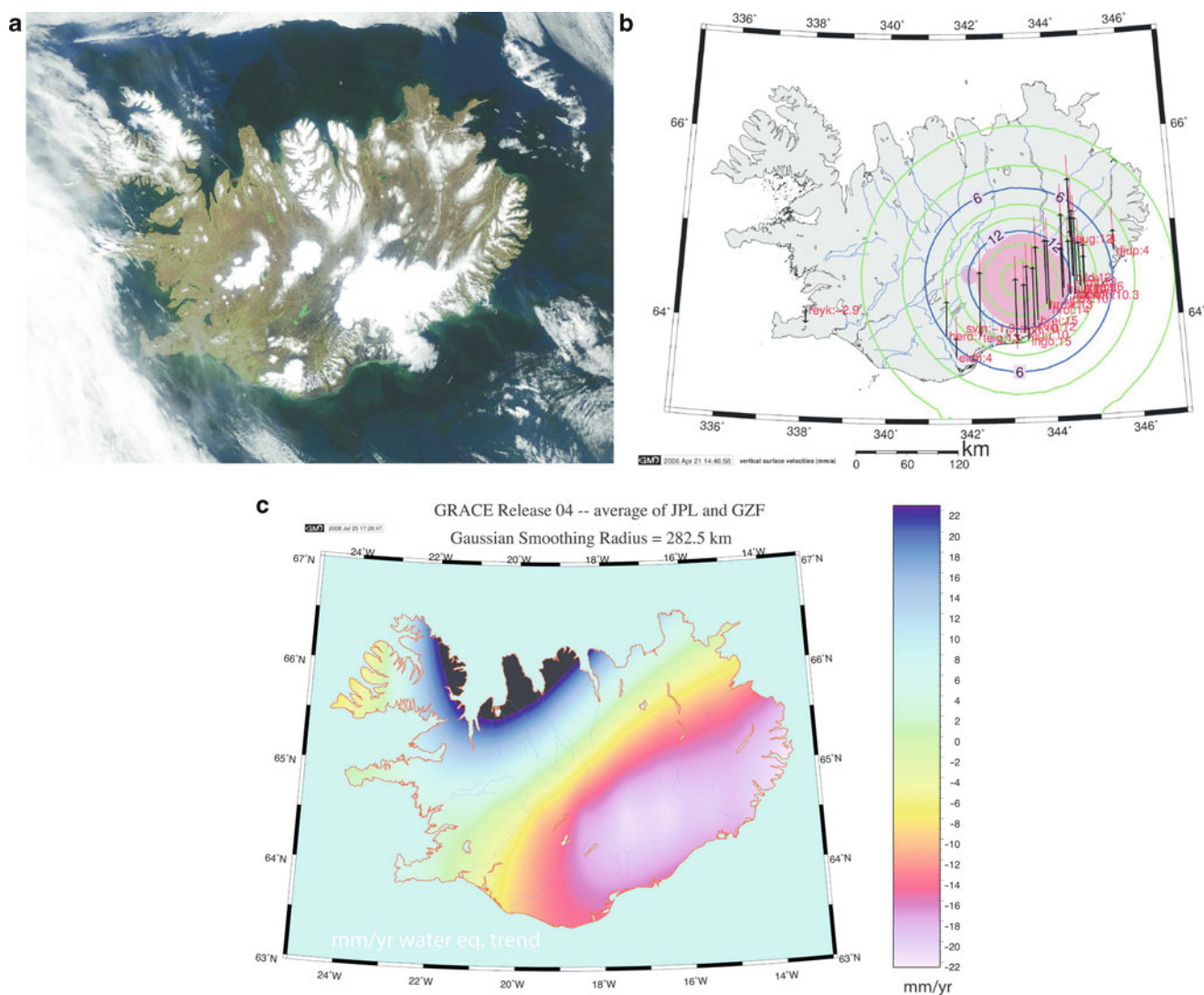
environments such as Alaska (Luthcke et al., 2008) with resolution approaching the inter-satellite tracking separation distance (~ 240 km).

To illustrate the importance of the cryospheric measurements, adding all of the mass imbalance estimates recovered from GRACE alone since its launch in 2002, including that of Chen et al. (2007) for southern Patagonia, these add to nearly 1.7 mm/year equivalent uniform global sea-level rise, nearly half that observed for the last 15 years (Church and White, 2006) and possibly the lion's share of that inferred to be caused by all continental sources of water mass imbalance. What then is the role of monitoring solid Earth related gravity change?

Mantle deformation

To monitor the hydrological or cryospheric mass changes using a time series constructed from individual GRACE maps, or equivalently harmonic coefficients, $C_{nm}(t_i)$, a coefficient representing water mass must be constructed, $^{WE}C_{nm}(t_i)$, by finding the deviation at time t_i from the average of the sum of maps at all times, t_i . In the presence of surface water masses, the instantaneous elastic deformation of the solid Earth can be accounted for using theory for loading of a spherically symmetric radially stratified Earth model (e.g., Farrell, 1972) and an n -dependent factor $h_n^e/(1 + k_n^e)$, where h_n^e and k_n^e are the radial displacement and potential Love numbers, respectively ($h_n^e \leq -1.01$, $-0.3 < k_n^e < 0$, for all $n > 1$) (e.g., Wahr et al., 1998). For example, if surface water mass builds in a region, it depresses the solid surface and the radially stratified density structure of the entire Earth; thus, the accelerations on the satellite are *less* than those that would correspond the same water mass on a *rigid* Earth. These elastic corrections for solid Earth deformation are routine and straightforward. The main complexities arise when GIA-related solid Earth motion and gravitational potential changes are locally, or regionally, coincident with the location of the water mass changes. Such is the case in Antarctica, as has been discussed by Velicogna and Wahr (2006b) and Ramillien et al. (2006). However, the coincidence of regional changes in hydrology can also be difficult to separate from GIA, where GIA dominates the secular gravity changes, such as in central Canada and central Scandinavia. Ivins and Wolf (2008) have recently reviewed the current status of overlaps in GIA and hydrology.

The ice caps and glaciers of Iceland provide an excellent example of how this coincidence of viscoelastic GIA and ice mass change-related gravity changes are measured. There is now consensus that such ice masses are melting and raising global sea level at an accelerating pace (e.g., Kaser et al., 2006). The main center of cryospheric mass in Iceland is located in the southeastern part of the island, as seen in the cloud-free Terra Satellite image of Figure 3a, where ice- and snow-dominated regions are white in color. A large thermal plume, probably rising from the core-mantle boundary (e.g., Nataf, 2000),



Solid Earth Mass Transport, Figure 3 Iceland true color image in summer using the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument onboard the Terra spacecraft (a) (Credit: http://visibleearth.nasa.gov/view_rec.php?id=6606). Vatnajökull ice cap is the larger white area in the southeast. GRACE monthly solution trends from release 04 (February 2003–April 2007) using an isotropic Gaussian smoothing filter applied to the harmonics with radius $\alpha = 282.5$ km. The units are in millimeter of water-height equivalent (b). Solid Earth uplift predictions at present day on an incompressible viscoelastic mantle half-space of viscosity 3×10^{18} Pa s under an elastic lithosphere of 25 km thickness. The ice cap evolves in accordance with the history simulated by Marshall et al. (2005) and calibrated by regional paleoclimate indicators (e.g., Flowers et al., 2008). This is a simple two-disk model of the Little Ice Age evolution of the Vatnajökull ice cap that evolves into the late twentieth century with the same observational constraints as in Fleming et al. (2007). GPS data are from UNAVCO archives and *up arrows* indicate uplift and down subsidence (with formal error estimates in red). These data are from a combination of permanent and episodic GPS tracking stations, references to International Terrestrial Reference Frame ITRF2000 on bedrock and have more than 6 years of observing history.

characterizes the regional mantle. A greatly reduced regional upper mantle viscosity is generally accepted, likely at values close to 3×10^{18} Pa s (Árnadóttir et al., 2005). This viscosity reduction greatly exacerbates both the prediction of the *sensitivity to*, and the *amplitude of*, GIA-related geodetic quantities. Icelandic paleoclimate and glacier history are relatively well known (e.g., Marshall et al., 2005; Flowers et al., 2008), as are recent mass

changes of the Vatnajökull ice cap and other major Icelandic glaciers (Magnússon et al., 2005). It is possible, therefore, to use a relatively complete ice load history and a parameter search of mantle viscoelastic properties to match the uplift patterns and magnitudes measured by GPS adjacent to the Vatnajökull ice cap as shown in Figure 3b. The model used to compute uplift in Figure 3b assumes an idealized circular disk to represent the ice cap

with the computational methods described by Ivins and James (1999) and includes the present-day responses to current ice mass changes (for a detailed description of viscoelastic GIA modeling of Iceland, readers should refer to recent work by Fleming et al., 2007). Ignoring the ice loss, the viscoelastic uplift of bedrock, $\partial_t u_b$, in response to ice mass loss (Figure 3b), the water-equivalent rate of roughly $\rho_{MC}/\rho_w \times \partial_t u_b$, where ρ_{MC}/ρ_w is the ratio of bedrock to water density, or about 2.8–3.8. The negative water-height rate deduced from GRACE spherical harmonic release 04 is in excess of 20 mm/year and has a local minima at the geographical position of the Vatnajökull ice cap (Figure 3c). Due to the unique circumstance of having GPS uplift data adjacent to the ice cap, and a well-constrained ice history, when the GIA phenomenon is correctly accounted for, the ice mass loss rate solution for eastern Iceland is roughly 5–7 Gt/year and would be almost half this value if GIA-related mantle mass variability had not been accounted for. This example of the Vatnajökull ice cap is for illustrative purposes, as the GRACE errors are quite poorly calibrated at this wavelength. The scales over which this GRACE rate is extracted from harmonic solutions are at the upper limits of resolution, shorter, for example, than the 300–400 km resolution in the example of mantle deformation and mass movement from the coseismic offsets caused by the M_w 9.1 December 2004 Aceh-Andaman earthquake (Han et al., 2006) and from Patagonian ice mass loss (Chen et al., 2007).

Conclusion

Future space-gravimetry systems, much like the highly successful polar-orbiting GRACE satellite pair, strive to meet a critical environmental remote-sensing challenge of the twenty-first century: determination of the long-wavelength mass stability of the Earth's continental fresh water supply (e.g., Velicogna and Wahr, 2006a, b; Ramillien et al., 2006; Syed et al., 2008). Terrestrial remote-sensing data have recently been directed at the severe consequences of the unprecedented warmth that has appeared during the last 100 hundred years and that now appears to be accelerating to paces unprecedented during the past 2,000 years (Mann, 2007). Monitoring the secular trends in both continental water storage and sustained ice sheet change will be of increasing societal concern. Use of space-determined gravity fields is strongly influenced by viscoelastic motions of the mantle and crust, and the study of such mass variability deep in the Earth's interior will be necessary for fully retrieving accurate and well-resolved trends in changing cryosphere and surface hydrology.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Árnadóttir, T., Jónsson, S., Pollitz, F. F., Jiang, W., and Feigl, K. L., 2005. Postseismic deformation following the June 2000 earthquake sequence in the south Iceland seismic zone. *Journal of Geophysical Research*, **110**, B12308, doi:10.1029/2005JB003701.
- Chen, J. L., Wilson, C. R., Tapley, B. D., Blankenship, D. D., and Ivins, E. R., 2007. Patagonian icefield melting observed by Gravity Recovery and Climate Experiment (GRACE). *Geophysical Research Letters*, **34**, L22501, doi:10.1029/2007GL031871.
- Cheng, M. K., Shum, C. K., and Tapley, B. D., 1997. Determination of long-term changes in the Earth's gravity field from satellite laser ranging observations. *Journal of Geophysical Research*, **102**, 22377–22390.
- Church, J. A., and White, N. J., 2006. A 20th century acceleration in global sea-level rise. *Geophysical Research Letters*, **33**, L01602, doi:10.1029/2005GL024826.
- Ciufolini, I., and Pavlis, E. C., 2004. A confirmation of the general relativistic prediction of the Lense-Thirring effect. *Nature*, **431**, 958–960.
- Cox, C. M., Klosko, S. M., and Chao, B. F., 2001. Changes in ice-mass balance inferred from the time variations of the geopotential observed through SLR and DORIS tracking. In Sideris, M. (ed.), *Gravity, Geoid and Geodynamics 2000*. Berlin: Springer. IAG Symposia Series, Vol. 123, pp. 355–360.
- Davis, J. L., Tamisiea, M. E., Elósegui, P., Mitrovica, J. X., and Hill, E. M., 2008. A statistical filtering approach for Gravity Recovery and Climate Experiment (GRACE) gravity data. *Journal of Geophysical Research*, **113**, B04410, doi:10.1029/2007JB005043.
- Farrell, W. E., 1972. Deformation of the Earth by surface loads. *Reviews of Geophysics and Space Physics*, **10**, 761–797.
- Fleming, K., Martinec, Z., and Wolf, D., 2007. Glacial-isostatic adjustment and the viscosity structure underlying the Vatnajökull ice cap. *Iceland, Pure and Applied Geophysics*, **164**, 751–768.
- Flowers, G. E., Björnsson, H., Geirsdóttir, A., Miller, G. H., Black, J. L., and Clarke, G. K. C., 2008. Holocene climate conditions and glacier variation in central Iceland from physical modeling and empirical evidence. *Quaternary Science Reviews*, **27**, 797–813.
- Forte, A. M., and Woodward, R. L., 1997. Seismic-geodynamic constraints on vertical flow between the upper and lower mantle: the dynamics of the 670 km seismic discontinuity, Chapter 10. In Crossley, D. J. (ed.), *The Earth's Deep Interior: The Dobboos Memorial Volume*. Amsterdam: Gordon and Breach Science, pp. 337–404.
- Han, S.-C., Shum, C. K., Bevis, M., Ji, C., and Kuo, C.-Y., 2006. Crustal dilatation observed by GRACE after the 2004 Sumatra-Andaman earthquake. *Science*, **313**, 658–662.
- Hughes, C. W., and Stepanov, V. N., 2004. Ocean dynamics associated with rapid J2 fluctuations: importance of circumpolar modes and identification of a coherent Arctic mode. *Journal of Geophysical Research*, **109**, C06002, doi:10.1029/2003JC002176.
- Ivins, E. R., and James, T. S., 1999. Simple models for late Holocene and present-day Patagonian glacier fluctuations and predictions of a geodetically detectable isostatic response. *Geophysical Journal International*, **138**, 601–624.
- Ivins, E. R., and Wolf, D., 2008. Glacial isostatic adjustment: new developments from advanced observing systems and modeling. *Journal of Geodynamics*, doi:10.1016/j.jog.2008.06.002.
- James, T. S., and Ivins, E. R., 1997. Global geodetic signatures of the Antarctic ice sheet. *Journal of Geophysical Research*, **102**, 605–633.

- Kaser, G., Cogley, J. G., Dyurgerov, M. B., Meier, M. F., and Ohmura, A., 2006. Mass balance of glaciers and ice caps: consensus estimates for 1961–2004. *Geophysical Research Letters*, **33**, L19501, doi:10.1029/2006GL027511.
- Kaula, W. M., 1983. Inferences of variations in the gravity field from satellite-to-satellite range rate. *Journal of Geophysical Research*, **88**, 8345–8349.
- Kozai, Y., 1968. Love's number of the Earth derived from satellite observation. *Publications of the Astronomical Society of Japan*, **20**, 24–26.
- Luthcke, S., Arndt, A. A., Rowlands, D. D., McCarthy, J. J., and Larsen, C. F., 2008. Recent glacier mass changes in the Gulf of Alaska region from GRACE mascon solutions. *Journal of Glaciology*, **54**, 767–777.
- Magnússon, E., Björnsson, H., Dall, J., and Pálsson, F., 2005. Volume changes of Vatnajökull ice cap, Iceland, due to surface mass balance, ice flow, and subglacial melting at geothermal areas. *Geophysical Research Letters*, **32**, L05504, doi:10.1029/2004GL021615.
- Mann, M. E., 2007. Climate over the past two millennia. *Annual Review of Earth and Planetary Sciences*, **35**, 111–136, doi:10.1146/annurev.earth.35.031306.140042.
- Marshall, S. J., Björnsson, H., Flowers, G. E., and Clarke, G. K. C., 2005. Simulation of Vatnajökull ice cap dynamics. *Journal of Geophysical Research*, **110**, F03009, doi:10.1029/2004JF000262.
- Muller, P., and Sjogren, W., 1968. Mascons: lunar mass concentrations. *Science*, **161**, 680–684.
- Nataf, H.-C., 2000. Seismic imaging of mantle plumes. *Annual Review of Earth and Planetary Sciences*, **28**, 391–417, doi:10.1146/annurev.earth.28.1.391.
- Paddack, S. J., 1967. On the angular momentum of the solid Earth. *Journal of Geophysical Research*, **72**, 5760–5762.
- Peltier, W. R., 1982. Dynamics of the ice age earth. *Advances in Geophysics*, **24**, 1–146.
- Petrovskaya, M. S., Nersisyan, A. N., and Pavlis, N. K., 2001. New analytical and numerical approaches for geopotential modeling. *Journal of Geodesy*, **75**, 661–672, doi:10.1007/s001900100215.
- Phillips, R. J., and Lambeck, K., 1980. Gravity fields of the terrestrial planets – long-wavelength anomalies and tectonics. *Reviews of Geophysics and Space Physics*, **18**, 27–76.
- Ramillien, G., Lombard, A., Cazenave, A., Ivins, E. R., Remy, F., and Biancale, R., 2006. Interannual variations of the mass balance of the Antarctic and Greenland ice sheets from GRACE. *Global and Planetary Change*, **53**, 198–208.
- Rapp, R. H., and Pavlis, N. K., 1990. The development and analysis of geopotential coefficient models to spherical harmonic degree 360. *Journal of Geophysical Research*, **95**, 21885–21911.
- Reigber, C., Balmino, G., Schwintzer, P., Biancale, R., Bode, A., Lemoine, J. M., Konig, R., Loyer, S., Neumayer, H., Marty, J. C., Barthelmes, F., Perosanz, F., and Zhu, S. Y., 2003. Global gravity field recovery using solely GPS tracking and accelerometer data from CHAMP. *Space Science Reviews*, **108**, 55–66.
- Richards, M. A., and Hager, B. H., 2004. Geoid anomalies in a dynamic Earth. *Journal of Geophysical Research*, **89**, 5987–6002.
- Rubincam, D. P., 1984. Postglacial rebound by Lageos and the effective viscosity of the lower mantle. *Journal of Geophysical Research*, **89**, 1077–1087.
- Rummel, R., Rothacher, M., and Buetler, G., 2005. Integrated Global Geodetic Observing System (IGGOS) – science rationale. *Journal of Geodynamics*, **40**, 357–362.
- Syed, T. H., Famiglietti, J. S., Rodell, M., Chen, J., and Wilson, C. R., 2008. Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research*, **44**, W02433, doi:10.1029/2006WR005779.
- Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M. M., 2004. GRACE measurements of mass variability in the Earth system. *Science*, **305**, 503–505.
- Tosi, N., Sabadini, R., Marotta, A. M., and Vermeersen, L. L. A., 2005. Simultaneous inversion for the Earth's mantle viscosity and ice mass imbalance in Antarctica and Greenland. *Journal of Geophysical Research*, **110**, doi:10.1029/2004JB003236.
- Velicogna, I., and Wahr, J., 2006a. Acceleration of Greenland ice mass loss in spring 2004. *Nature*, **443**, 329–331.
- Velicogna, I., and Wahr, J., 2006b. Measurements of time-variable gravity show mass loss in Antarctica. *Science*, **311**, 1754–1756.
- Wahr, J., Molenaar, M., and Bryan, F., 1998. Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE. *Journal of Geophysical Research*, **103**, 30205–30230.
- Yoder, C. F., Williams, J. G., Dickey, J. O., Schutz, B. E., Eanes, R. J., and Tapley, B. D., 1982. J2 from LAGEOS and the non-tidal acceleration of Earth rotation. *Nature*, **303**, 757–762.

Cross-references

[Cryosphere, Climate Change Effects](#)
[Cryosphere, Measurements and Applications](#)
[Global Earth Observation System of Systems \(GEOSS\)](#)
[Sea Level Rise](#)

STRATOSPHERIC OZONE

Michelle Santee

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition

Stratosphere. The region of the atmosphere extending from about 13 to 50 km altitude.

Ozone. A naturally occurring trace gas in Earth's atmosphere. An ozone molecule consists of three atoms of oxygen bound together in a triangular arrangement.

Introduction

The vast majority of ozone (about 90 %) resides in the stratosphere, where it plays a crucial role in protecting life on Earth from harmful solar ultraviolet radiation. Most of the remaining ozone is found in the troposphere, the part of the atmosphere between the surface and the stratosphere and the region in which most clouds and weather occur. Tropospheric ozone is an important pollutant and a greenhouse gas. In addition to its direct influence on air quality, tropospheric ozone helps to regulate the ability of the atmosphere to “cleanse” itself of many other pollutants. Anthropogenic (human-induced) increases in ozone at ground level, where it is a key component of photochemical smog, can have very deleterious effects. Thus, although all ozone molecules are chemically identical, their environmental impacts depend strongly on where they are situated in the atmosphere. In this entry, we concentrate on ozone in the stratosphere. In recent decades, anthropogenic emissions have severely compromised the stability of the [stratospheric ozone](#) layer. Stratospheric ozone depletion is a global problem that has its most dramatic manifestation in the polar regions,

especially over Antarctica, where an “ozone hole” has formed each spring for the last few decades.

Ozone production and distribution

The abundance of ozone is relatively low, with peak mixing ratios (the ratio of the concentration of ozone to that of air) of only about 10–12 parts per million by volume (ppmv) (e.g., McPeters et al., 2007). Ozone is commonly measured in terms of its column abundance, the vertically integrated density overhead at a given location. Total ozone column abundances are traditionally reported in Dobson units (DU, named for G.M.B. Dobson, who in the 1920s pioneered routine measurements of ozone column amounts), which correspond to the thickness in millimeters that the ozone column would have if all the gas were at standard temperature and pressure. Typical ozone column abundances range from 200 to 500 DU, depending on the latitude and season (e.g., Dessler, 2000).

Despite its low concentration, ozone is arguably the most important stratospheric constituent. It is an extremely efficient absorber in the ultraviolet (UV); as a result, an intact ozone layer prevents practically all solar radiation at wavelengths shorter than 300 nm from reaching the Earth's surface. Excessive amounts of UV-B (280–315 nm) radiation are particularly injurious to humans and other animals, terrestrial and aquatic ecosystems, and physical materials. It is therefore not an exaggeration to say that without Earth's protective ozone shield, life as we know it could not have evolved.

The distribution of stratospheric ozone is controlled by the interplay between chemistry and transport by the prevailing winds. Ozone is produced in the stratosphere when molecular oxygen (O_2) is photolyzed (i.e., split apart) by solar UV radiation. The resulting two oxygen atoms (O) each rapidly combine with other oxygen molecules to produce two ozone molecules (O_3). The overall result is that, in the presence of sunlight, three oxygen molecules react to form two ozone molecules. Because these reactions take place wherever UV radiation is available, the greatest ozone production occurs in the tropical stratosphere. Interestingly, however, the largest total column ozone amounts are found at middle and high latitudes in late winter and early spring (after several months of total or near-total darkness), whereas the tropics actually contain the smallest column ozone amounts. This lack of correlation between the regions of highest production and those of highest abundance underscores the importance of transport in determining the overall morphology of stratospheric ozone. Winds circulate air in the stratosphere, with rising motion in the tropics, poleward motion aloft, and sinking motion in the polar regions. Known as the Brewer-Dobson circulation, this flow carries air rich in ozone away from the tropics and toward the poles.

Although ozone molecules are continuously produced, they are also continuously destroyed in chemical reactions

with oxygen atoms and oxides of nitrogen, hydrogen, chlorine, and bromine. These reactions are termed “catalytic” because the species that initiate ozone destruction are not themselves consumed in the process, allowing the cycle to repeat many times. The net effect of the catalytic cycle is the conversion of O_3 molecules into O_2 molecules. The catalysts in these reactions are free radicals (i.e., molecules with an unpaired electron, making them highly reactive). The radicals are derived from long-lived species (which in most cases have both natural and anthropogenic sources) transported from the troposphere into the stratosphere by the large-scale circulation. Typical mixing ratios of the oxides of nitrogen are in the part per billion (ppbv) range, while those of the reactive hydrogen, chlorine, and bromine species are even lower. Thus, ozone abundances are in large part regulated by radical species present in the stratosphere in even more minute quantities.

The atmosphere normally maintains a balance, or steady state, between photochemical production and catalytic destruction of ozone, with the total level in the stratosphere remaining fairly constant. Small variations in total ozone occur because of seasonal changes in the strength of atmospheric transport, the intensity of incident sunlight, stratospheric temperatures, and other factors. Natural variations can arise from other sources as well; for example, fluctuations in UV radiation over the course of the 11 year solar cycle alter ozone production rates slightly (1 – 2 %) (e.g., Brasseur and Solomon, 2005).

Chlorine-catalyzed ozone depletion in the polar lower stratosphere

Over the last several decades, the natural balance between ozone formation and destruction has been perturbed as manufactured chemicals have entered the stratosphere. It is now understood that the severe springtime depletion of the [stratospheric ozone](#) layer over Antarctica – the so-called ozone hole – is caused by chlorine and bromine chemistry (known collectively as halogen chemistry) (e.g., Solomon, 1999; Fahey, 2007). The primary source of stratospheric chlorine is chlorofluorocarbons (CFCs), chemical compounds composed of chlorine, fluorine, and carbon. CFCs were introduced in the 1930s and used in a variety of industrial and commercial applications, including as refrigerants, solvents, and aerosol propellants, as well as in foam packaging. CFCs were used so extensively because they have low boiling points and are stable, nonflammable, and unreactive in the lower atmosphere. Other halogen source gases include carbon tetrachloride, methyl chloride, and methyl chloroform, as well as methyl bromide and the bromine-containing “halons” used as fire retardants.

Chlorine and bromine are also released at ground level through other human activities and natural processes. Common examples include emission of chlorine gases from swimming pools, wastewater treatment, and household bleach, as well as from volcanic ejecta; chlorine is also present in sea salt produced by evaporation of

ocean spray. In all of these cases, however, the chlorine is in forms that are highly reactive or soluble and are thus washed out in rain or ice in the lower regions of the atmosphere before they can be delivered to the stratosphere. In contrast, the very properties that made CFCs so useful turned out to pose a grave environmental threat. Because they are inert and insoluble in water, CFCs spread throughout the troposphere in a matter of months and are then carried along as the overturning circulation lofts air into the upper stratosphere. There, above the protection afforded by the bulk of the ozone layer, CFCs are broken down by high-energy solar UV radiation, liberating chlorine atoms. The transit time between the troposphere and the stratosphere is about 5 years. During each circuit through the Brewer-Dobson circulation, only about 10 % of the mass of the troposphere is exchanged with the upper stratosphere, and only a fraction of CFC molecules is decomposed; thus, most CFCs have long atmospheric lifetimes (about 50–500 years) (e.g., Brasseur and Solomon, 2005).

In the mid-1970s, it was recognized that chlorine released from CFCs could destroy stratospheric ozone in a catalytic cycle (Molina and Rowland, 1974), a notion that generated enormous scientific interest. For their work in helping to elucidate the sensitivity of the ozone layer to anthropogenic emissions, Mario Molina and F. Sherwood Rowland were awarded (along with Paul Crutzen) the 1995 Nobel Prize in Chemistry. Even after the connection with chlorine from CFCs was made, however, model calculations indicated that ozone loss would be gradual and moderate (5 – 20 %), limited mainly to altitudes in the upper stratosphere (near 40 km) above the peak in the ozone profile. Consequently, scientists and the public alike were shocked by the news in 1985 that ground-based measurements from the British Antarctic Survey station at Halley Bay, Antarctica, had revealed a dramatic downturn in October average column ozone over the previous few years (Farman et al., 1985). NASA satellite measurements subsequently confirmed the rapid decline in total ozone and demonstrated that the phenomenon extended over a vast geographic region roughly encompassing the Antarctic continent (Stolarski et al., 1986). Vertical profile measurements from balloon-borne ozonesondes showed the loss to be largely confined to the lower stratosphere, with virtually all ozone in the layer between about 14 and 20 km removed within a period of 4–6 weeks (Hofmann et al., 1987; Solomon, 1999; Fahey, 2007). Such severe loss in the lower stratosphere, where ozone mixing ratios are normally largest, leads to a reduction in total column ozone of as much as 60 – 70 %.

Early models failed to predict the development of the ozone hole because they focused exclusively on homogeneous (gas-phase) chemistry. A series of coordinated field campaigns to Antarctica, together with laboratory experiments and modeling studies, provided the information necessary to establish the cause of the massive ozone loss within a few years of its discovery. The first step is the formation of the “polar vortex.” In the continuous

darkness of winter polar night, temperatures drop very low, leading to a large temperature gradient near the polar terminator (the line delimiting the unilluminated region), which in turn induces a steep pressure gradient. The resulting flow is deflected by the Coriolis force, creating a band of intense westerly (i.e., eastward) winds encircling the pole. This strong wind jet provides a barrier to mixing, effectively isolating the region poleward of it, termed the polar vortex, from lower-latitude air. In the very low temperatures inside the lower stratospheric winter polar vortex, water vapor (H_2O) and nitric acid (HNO_3) condense to form polar stratospheric clouds (PSCs). PSCs were long known to occur – for example, they were described in 1911 by noted Antarctic explorer Robert Falcon Scott and his party – but they remained little more than scientific curiosities until the ozone hole was reported.

PSCs are a crucial link in the chain leading to severe ozone destruction. Their particles provide surfaces on which heterogeneous reactions (i.e., reactions in which one reactant is absorbed onto/into the particle while the other remains in the gas phase) can take place very rapidly. These reactions convert chlorine from relatively benign “reservoir” species such as hydrogen chloride (HCl) and chlorine nitrate (ClONO_2) to intermediate forms such as molecular chlorine (Cl_2) and hypochlorous acid (HOCl) that are quickly photolyzed. In this manner, vortex air is primed for ozone destruction during the cold, dark winter months. When sunlight returns to the polar regions in spring, the intermediate species are rapidly broken apart to produce highly reactive ozone-destroying forms. The predominant form of reactive chlorine in the stratosphere – the “smoking gun” that signals chlorine-catalyzed ozone destruction – is the chlorine monoxide radical, ClO .

In the mid-1980s, two sets of measurements cemented the link between elevated levels of ClO and severe ozone destruction in the lower stratosphere. First, ground-based microwave emission measurements from McMurdo Station, Antarctica, indicated greatly enhanced ClO near 20 km, with mixing ratios of more than 1 ppbv, as much as two orders of magnitude larger than predicted, based on standard gas-phase photochemistry (de Zafra et al., 1987). Second, in situ measurements from the NASA ER-2 high-altitude research aircraft (a converted U2 spy plane) revealed an abrupt increase in ClO closely correlated with a steep decline in ozone along the flight track inside the Antarctic polar vortex (Anderson et al., 1989). The large enhancements in ClO found by these studies could only have arisen through the near-total transformation of chlorine from reservoir to reactive species, and the spatial and temporal variation of the ClO/O_3 anticorrelation confirmed that ozone loss was indeed driven by chlorine chemistry. It was not until the launch of the Microwave Limb Sounder (MLS) on NASA's Upper Atmosphere Research Satellite (UARS) in September 1991 that researchers were able to map the full three-dimensional distribution of ClO in the stratosphere and

track its seasonal evolution around the globe. UARS MLS found an almost exact coincidence between enhanced ClO and depleted ozone throughout the Antarctic vortex (Waters et al., 1993).

UARS MLS measurements also verified that ClO enhancement in the Arctic polar vortex can be comparable, in terms of both magnitude and areal extent, to that in the Antarctic. Nevertheless, the Arctic has so far been spared an annual ozone hole. This interhemispheric asymmetry in the severity of ozone destruction is attributable to differences in meteorological conditions that arise in large part because of differences in the distribution of oceans and continents, in particular mountains at middle and high latitudes. Winds flowing over mountains generate waves that propagate into the stratosphere and disturb the polar vortex. As a result, the Arctic vortex exhibits much greater interannual variability than its southern counterpart and is typically warmer (by 10–20 K), weaker, more permeable, smaller, and shorter lived, leading to fewer, less persistent PSCs. In the cold, isolated Antarctic vortex, PSC particles grow sufficiently large that they undergo appreciable gravitational sedimentation. As they fall, they irreversibly remove from the stratosphere the HNO_3 (and H_2O) sequestered in them in a process known as denitrification (dehydration). Unlike the Antarctic, the Arctic generally experiences minimal denitrification. Photolysis of gas-phase HNO_3 in springtime produces nitrogen dioxide (NO_2), which combines with ClO, to reform the reservoir ClONO_2 . This reaction is the primary pathway for chlorine deactivation in the Arctic. For all of these reasons, the horizontal and vertical extent, duration, and magnitude of ClO enhancement are typically much smaller in the Arctic than in the Antarctic, and ozone loss, although it does occur in many winters, is correspondingly smaller.

Midlatitude ozone depletion

Though much less dramatic than those observed in the polar regions, statistically significant downward trends have also been detected in ozone at middle latitudes (e.g., Solomon, 1999). A substantial fraction of the midlatitude depletion is attributable to dilution with processed air exported from the decaying vortex at the end of winter, especially in the Southern Hemisphere, where polar ozone loss is severe. Another factor is in situ chemical destruction. Heterogeneous reactions akin to those that activate chlorine on PSCs also take place (albeit more slowly) on sulfate aerosol droplets, which are ubiquitous throughout the lower stratosphere even under background conditions. Major volcanic eruptions, which inject large amounts of sulfur gases directly into the stratosphere, can greatly enhance stratospheric aerosol surface area density and intensify ozone loss. In fact, substantial decreases in column ozone (on the order of 5–10%) were recorded over large regions of the globe following the eruption of Mt. Pinatubo in the Philippines in June 1991 (Solomon, 1999; Dessler, 2000). It is important to note, however, that the eruption of Mt. Pinatubo would

not have had a significant impact on global ozone had stratospheric chlorine loading not far exceeded natural levels.

Ozone recovery and climate change

At its peak around the year 2000 (Fahey, 2007), stratospheric chlorine reached levels more than six times its natural abundance. Since then, however, chlorine has been declining in response to regulations enacted under an international agreement known as the *Montreal Protocol on Substances that Deplete the Ozone Layer*. The 1987 Protocol, and subsequent more stringent Amendments and Adjustments, established legally binding controls on the production of halogen source gases. Eventually ratified by over 190 nations, these provisions averted a global environmental calamity. The stratospheric ozone layer has now stabilized, but more than 20 years after the signing of the original Protocol, large ozone holes continue to occur, attaining an areal extent of about 25 million km^2 (roughly twice the surface area of the Antarctic continent) in an average year. Because of the long residence times of CFCs, it will take several more decades to cleanse the stratosphere of excess chlorine and curtail ozone depletion. Assuming continued worldwide compliance with the Protocol, stratospheric chlorine and bromine loading are expected to return to pre-1980 values by about 2050 (Fahey, 2007).

Recovery of the ozone layer does not depend solely on diminishing anthropogenic halogen abundances, however. Earth's changing climate may have a significant effect: While the surface of the planet is warming in response to increasing greenhouse gases, the stratosphere is cooling. PSC lifetimes are prolonged in a colder lower stratosphere, exacerbating polar ozone depletion. This could delay recovery and trigger development of an ozone hole in the Arctic, where wintertime temperatures often hover near PSC formation thresholds. But assessment of the sensitivity of ozone chemistry to temperature is complicated by concomitant changes in dynamics and transport. Greenhouse-gas-induced acceleration of the Brewer-Dobson circulation is expected to alter the concentrations of water vapor and trace gases in the stratosphere, which in turn influence ozone. Moreover, lower temperatures will slow catalytic ozone destruction cycles outside of the polar lower stratosphere, hastening recovery in those regions. On the other hand, higher surface temperatures may modify emission rates of some naturally occurring halogen source gases. Increased abundances of other pollutants arising from human activities may also perturb the balance between global ozone production and loss. Separating the effects of these competing factors is challenging, making future changes in ozone difficult to predict. Current state-of-the-art models show a range of projections, with most indicating full recovery around the middle of the twenty-first century at middle latitudes and slightly later in the Antarctic (Fahey, 2007). It is important to note, however, that because of changes in

climate and other parameters, the atmosphere is unlikely to return to a state identical to that before the onset of anthropogenic ozone depletion.

Conclusion

The last few decades have seen enormous changes in the stratospheric ozone layer so vital to life on Earth. The ozone hole serves as a cautionary tale of how an apparently harmless class of chemical compounds can prove to have disastrous consequences and thus highlights how human activities can change the natural state of our atmosphere in unintended ways. But it is also a story of hope that shows how the scientific community, national governments, and industry can identify a global environmental threat and work together to mitigate it. It is imperative that we as a society continue to monitor the stability of the ozone shield to see how the story ends in light of another environmental challenge: anthropogenic climate change.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Anderson, J. G., Brune, W. H., and Proffitt, M. H., 1989. Ozone destruction by chlorine radicals within the Antarctic vortex: the spatial and temporal evolution of ClO – O₃ anticorrelation based on in situ ER-2 data. *Journal of Geophysical Research*, **94**, 11465.
- Brasseur, G., and Solomon, S., 2005. *Aeronomy of the Middle Atmosphere*, 3rd revised and enlarged edn. Dordrecht: Springer.
- de Zafra, R. L., Jaramillo, M., Parrish, A., Solomon, P., Connor, B., and Barrett, J., 1987. High concentrations of chlorine monoxide at low latitudes in the Antarctic spring stratosphere: diurnal variation. *Nature*, **328**, 408.
- Dessler, A. E., 2000. *The Chemistry and Physics of Stratospheric Ozone*. San Diego: Academic.
- Fahey, D. W., 2007. *Twenty questions and answers about the ozone layer: 2006 update, scientific assessment of ozone depletion: 2006*. Geneva: World Meteorological Organization.
- Farman, J. C., Gardiner, B. G., and Shanklin, J. D., 1985. Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature*, **315**, 207.
- Hofmann, D. J., Harder, J. W., Rolf, S. R., and Rosen, J. M., 1987. Balloon-borne observations of the development and vertical structure of the Antarctic ozone hole in 1986. *Nature*, **326**, 59.
- McPeters, R. D., Labow, G. J., and Logan, J. A., 2007. Ozone climatological profiles for satellite retrieval algorithms. *Journal of Geophysical Research*, **112**, D05308, doi:10.1029/2005JD006823.
- Molina, M. J., and Rowland, F. S., 1974. Stratospheric sink for chlorofluoromethanes: chlorine atom catalyzed destruction of ozone. *Nature*, **249**, 810.
- Solomon, S., 1999. Stratospheric ozone depletion: a review of concepts and history. *Reviews of Geophysics*, **37**, 275.
- Stolarski, R. S., Krueger, A. J., Schoeberl, M. R., McPeters, R. D., Newman, P. A., and Alpert, J. C., 1986. Nimbus 7 satellite measurements of the springtime Antarctic ozone decrease. *Nature*, **322**, 808.

Waters, J. W., Froidevaux, L., Read, W. G., Manney, G. L., Elson, L. S., Flower, D. A., Jarnot, R. F., and Harwood, R. S., 1993. Stratospheric ClO and ozone from the microwave limb sounder on the upper atmosphere research satellite. *Nature*, **362**, 597.

Cross-references

Trace Gases, Stratosphere, and Mesosphere

SUBSIDENCE

Stuart Marsh¹ and Martin Culshaw^{2,3}

¹Nottingham Geospatial Institute, The University of Nottingham, Nottingham, UK

²Honorary Research Associate, British Geological Survey, Keyworth, Nottingham, UK

³Honorary Visiting Professor, School of Civil Engineering, University of Birmingham, Edgbaston, Birmingham, UK

Synonyms

Collapse; Compaction; Ground motion; Ground movement; Settlement

Definition

The term “subsidence” is often misunderstood and misused, both by specialist geoscientists and by the general public. As such, the term should be avoided and more precise terms used instead. Strictly, the term has been used to mean movement of the ground surface in a generally vertical, downward direction, without reference to the areal extent of movement, the rate at which movement takes place, or the cause of the movement. For example, Keller (1985) defined subsidence as “Sinking, settling, or other lowering of parts of the crust of the earth.” However, Johnson and DeGraff (1988) took some of the factors into account by defining subsidence as “. . . displacement of the ground surface vertically over a broad region or at localized areas. It may be either a gradual lowering or a collapse.” In the public’s mind, subsidence is sometimes associated with a particular cause, for example, mining subsidence, without an understanding that there are many other causes.

As remote sensing platform instrumentation can measure both subsidence (*sensu stricto*) and “negative” subsidence (rise), it is proposed, here, that the term “subsidence” should not be used and be replaced by the term “ground movement.” The definition of this term is proposed as: “The movement of the ground, however caused, in any direction, either vertically up or down, or laterally, or components of these. The areas involved can range from less than 1 m² to several square kilometers. The amount of movement can vary from a few millimeters to more than a kilometer. The length of time for the movement to be completed can vary from a few seconds to several years.”

The synonymous term “ground motion” is best reserved for the quantitative amount of ground movement observed and measured repeatedly by remote sensing platform instrumentation over a period of time. Hence, a rate of movement can be determined for different time periods. The movement measured is likely to be in the direction of the “line of sight” of the satellite instrumentation, which may not be vertical.

Importance of ground movement information

Ground movement causes death, injury, damage to buildings, structures, and infrastructure, and, hence, financial loss. It occurs on every continent and offshore. However, the amount of injury and loss depends upon the rate of ground movement and its magnitude. Sometimes, rapid and large ground movements will be preceded by slower and smaller movements that can be monitored. Examples of this include the growth of lava domes prior to an eruption and ground level rise/fall prior to an earthquake.

Even where ground movement takes place slowly and the amount of movement is only a few tens of millimeters, the financial cost can be great. In the United Kingdom (UK), damage to buildings caused by ground movement can be insured against, and in the last 20 years, losses have averaged between £3 and 400 million per year. Most of these losses were the result of the shrinkage of Mesozoic and Tertiary clay soils in southeast and central England. This does not represent the total loss due to ground movement as losses from ground movement caused by coal mining are not included in this total. (These losses are covered by a public body called the Coal Authority).

In France, where ground movement is also covered by insurance (though in a slightly different form), losses averaged over €300 million per year between 1989 and 2003. Again, the cause was the presence of expansive clay soils (Cheetham, 2008). Almost every country in the world has several geohazards that cause ground movement activity for at least part of the year.

To summarize, large-magnitude, rapid ground movements tend to cause both significant human and financial loss. On the other hand, small-magnitude, slow ground movements cause little human loss but can still result in substantial financial loss. As a result, the acquisition of ground movement information is very important if losses are to be anticipated. Research into the prediction of large-magnitude events has yet to identify ways of reliably predicting the location, magnitude, and timing of earthquakes and volcanic eruptions. However, frequent monitoring of ground movement, particularly using remote sensing, is likely to provide a major contribution to the anticipation of these events. Smaller and slower ground movements can also be monitored remotely, though the processes are better understood and hence are more predictable. Many of these processes are related to climatic conditions, and hence, prediction of large and/or intense rainfall events and drought is important. Predicted temperature rise is also important, for example, in relation to the thawing of thermokarst. In the future,

long-term monitoring of ground movement, however caused, is likely to become increasingly important in reducing losses arising from ground movement events.

Description of ground movement phenomena

Ground movement can have a wide range of natural and anthropogenic causes. Bell (2007) specifically referred to subsidence caused by mining, groundwater extraction, and dissolution of more soluble rocks, while Johnson and DeGraff (1988) listed dissolution, collapse of lava tube roofs, groundwater, oil or gas extraction, mining and swelling and shrinking clay soils as potential causes. Bell (1998) added fault movement to the identified causes and settlement as a low-rate and low-magnitude form of subsidence. To the lists of these authors can be added isostatic movement, seismic activity, volcanic activity, mass movement (including landsliding), soil fabric collapse, soil compressibility, and settlement of waste material.

Natural causes of ground movement

Most of the causes of ground movement discussed below are explained in detail in engineering, geological, and geomorphological textbooks (e.g., Bell, 1999; Keller et al., 2008; McCall et al., 1992; Nuhfer et al., 1993).

Earth tides: Movements of the Earth’s surface caused by the gravitational effects of the moon and the sun. These can take place diurnally or over longer periods. Movements are both vertical and horizontal and may amount to up to more than a 100 mm.

Isostatic movement

Readjustment to a state of equilibrium between the Earth’s lithosphere and asthenosphere. Lack of equilibrium is brought about by factors such as ice loading during glacial periods and movement of tectonic plates. Ground level rise from glacial unloading may amount to several meters over periods of thousands of years.

Seismic activity

While earthquakes, particularly large ones, can involve considerable and rapid ground movement (“strong ground motion”), usually associated with fault activity, slower- and longer-term ground movements may take place prior to, or as a result of, an earthquake. The “strong ground motions” associated with the earthquake occurrence may have no net ground movement effect. However, permanent uplift or depression of the ground surface over several square kilometers is known. Total change in ground level may amount to a meter, or more. For example, the Limón-Telire earthquake in Costa Rica in April 1991 caused between 0.3 and 1.9 m of uplift in the Limón area on the Caribbean coast.

Volcanic activity

The movement of volcanic magma toward the surface frequently leads to the uplift and subsidence of the area around an active volcano. Bell (1999) noted that the

summit of Kilauea, Hawaii, has risen by nearly a meter prior to eruption. However, such changes in ground level are not reliable indicators of eruption.

Dissolution

Rocks such as salt, gypsum, chalk, and limestone are susceptible to dissolution that can cause voids to be formed not far below the ground surface. If these collapse, ground movement is likely to result. The amount of movement depends upon the size of the void; the rate of movement (collapse) is likely to be high. Sometimes dissolution leads to the washing down of overlying soil particles into the dissolving rock, leading to gradual ground movement. The ground movement features are known as sinkholes or dolines and may range in diameter from a few meters to, exceptionally, more than 1,000 m (Waltham et al., 2005). Similar features can also be caused by the collapse of lava tubes where molten lava has run downslope leaving a solidified cover at the surface.

Mass movement

Landslides, flows, and falls involve the downward and lateral movement of soil, rock, or debris as a result of ground failure of a slope (Turner and Schuster, 1996). Movement can involve thousands of tonnes of material moving distances ranging from a few meters to, exceptionally, more than a kilometer.

Melting of permafrost (thermokarst)

As frozen ground thaws, settlement will occur as water is expelled by the overburden pressure and any artificial load. If the soil is free draining, the amount of settlement is likely to be greater. The amount of sinking of the ground movement will depend on the soils present and the rate of thawing. Heave can take place if the soil refreezes.

Shrinkage of peat

As peat dries out, it shrinks causing a lowering of the ground surface. This can amount to several meters. For example, in the Fens of eastern England, long-term drainage of the peat for agricultural purposes has caused a lowering of the ground surface of close to 10 m. Between 10 % and 75 % of the peat volume can be lost, and the change is permanent (Bell, 1999).

Swelling and shrinkage of clay soils

The main cause of swelling and shrinkage of clay soils is the presence of clay minerals that experience significant volume changes as their moisture content changes. Montmorillonite is a particularly susceptible clay mineral. Both shrinkage (causing downward movement of the ground surface) and swelling (causing upward movement) can result in damage to buildings. Amounts of movement range from a few millimeters to many tens of millimeters.

Soil consolidation

Soils consolidate (reduce in volume) due to overburden pressure and any artificially imposed load. This is

a process that takes place gradually as soils age and as natural processes ("weathering") act upon them. Under natural conditions, settlement will be slow and will amount to a few millimeters or tens of millimeters.

Soil fabric collapse

Some soils, particularly windblown silts and fine sands (loess), have a "metastable" microfabric caused by the way in which they were deposited. Quartz particles are separated by "bridges" of clay-sized minerals. The soils have high void ratios and low densities. On wetting, the microfabric collapses due to a reduction in the bonds between particles causing a reduction in void ratio and increase in density. The amount of volume change can range from 1 % to over 20 %. In extreme cases, settlements in excess of a meter have been observed (Bell, 1999).

Anthropogenic (artificial) causes of ground movement

Mining

Mining involves the removal of an economic solid mineral either at the ground surface (pits and quarries) or underground. Partial extraction methods involve leaving some of the mineral to support the roof or walls of the mine. However, over time, the supporting pillars may deteriorate and, ultimately, fail resulting in collapse of the mine workings and lowering of the ground surface. Total extraction methods involve the removal of almost all of the mineral and the controlled collapse of the overlying material into the void created. This results in lowering of the ground surface. The advantage is that the lowering of the ground surface is controlled and predictable. The amount of lowering may range from a few tens of millimeters to ten, or so, meters. Abandoned mine entrances (shafts and adits) may also fail creating either holes at the surface or its lowering.

Fluid withdrawal

Shrinkage of peat has been described above. However, subsidence, and consequent lowering of the ground surface, can take place when a fluid, usually groundwater, oil, or brine, is withdrawn from the ground. The overburden pressure is initially borne by the solid particles and the fluid in the pore space. As fluid is withdrawn, the load is transferred to the solid particles, increasing the so-called effective overburden pressure. This, in turn, causes consolidation of the sediment. The amount of subsidence depends upon the increase in effective pressure and the thickness and nature of the sedimentary deposit. In the San Joaquin Valley of California, USA, 9 m of subsidence occurred between 1925 and 1977 due to groundwater extraction for agriculture. Recharge of the fluid into the ground can reduce the rate of subsidence and may result in some recovery, with the ground surface rising. However, full recovery cannot be achieved.

Fault reactivation

Geological faults can be locations where ground movement caused by mining is concentrated. This reactivation

of the faults may take place during mining and may continue for years afterward. Movements along a fault greater than 2 m can occur, particularly where several seams have been extracted. Fault reactivation is more associated with shallow mine workings (Bell and Donnelly, 2006).

Settlement of waste material

Though modern waste disposal sites are usually carefully controlled, both in terms of their content and of how the site is constructed, older sites and those where controls are limited may contain mixtures of materials that have not been compacted. Over time, as the waste material alters chemically and as a result of the overburden pressure from dumping of later waste, the tip will settle, lowering the level of its surface. Amounts of settlement may range from millimeters to more than a meter, depending on the nature of the fill material and the degree of control of fluids entering the tip.

Engineering compaction

Compaction is a civil engineering process by which soil is densified by expelling air from the void space to achieve a stronger material. Usually, this involves artificially placed material (“fill”) used to construct embankments, dams, and other structures. The amount of lowering of the surface might range from millimeters to several tens of millimeters.

Measuring ground movement using remote sensing data

Ground movement can be measured using a variety of spaceborne, airborne, and ground-based remote sensing systems. Measuring ground movement requires a minimum of two data acquisitions over a study area: one to establish the baseline ground position and another to capture changes to it after a period of movement. The necessary time separation depends on a trade-off between the rate of movement and the ability of the sensor to resolve the ground movement. This measurement will capture the total ground movement in a specific time period. It can be extended via repeated data acquisitions to cover more time periods, leading to the concept of monitoring ground movement. Such monitoring allows us to measure ground motion and changes to its rate over time. The cost of monitoring is higher than that of measuring but may be justified by the impact of the ground movement on lives and economies. Ground movement is typically measured on a per-pixel basis and in the line of sight of the sensor. Of course, the ground movement may actually be vertical, horizontal, or a combination of both. There are various ways of accounting for the mismatch between the geometry of the movement and the viewing geometry of the sensor measuring it, to give a true measurement of ground movement in x , y , and z .

Spaceborne measurement techniques

Satellite altimetry

Altimeters are active remote sensing instruments. That is, they operate by sending a signal from the satellite to the surface under study and measuring how long it takes to return to the sensor. The signal may be radar or light from a laser. It is used to build up an image of the surface geometry. Changes to that geometry could, in principle, be measured via a second acquisition. However, the spatial resolution of such instruments is typically insufficient for application to ground movement. They are more commonly used in oceanography applications (Cheney et al., 1986).

Monoscopic optical satellites

It is possible to work out the horizontal component of ground movement using two passively acquired optical images. Once they are geometrically rectified and registered to each other precisely, image analysis software can be used to assess whether features in the first image, such as prominent outcrops, are in a different location in the second. This gives a simple measurement of horizontal ground movement. It is not widely used, because it cannot capture the vertical component and so has limited application in a 3D science like geology.

Stereoscopic optical satellites

Stereoscopic optical imagery from satellites like Landsat and SPOT can be used to produce a digital terrain model (DTM) with a resolution on the order of meters to tens of meters in a process known as photogrammetry (Toutin, 2001). The technique uses two images of the same area taken from adjacent orbits or two positions along track. These are compared using photogrammetric software, which takes advantage of parallax between the two viewing geometries to work out the surface geometry, in a similar way to the brain processing images from two eyes to generate stereoscopic vision. If the process is repeated, a second DTM results, and the two can be differenced to quantify changes in the DTM between acquisitions. These changes measure the ground movement. The main constraint on this technique is the need for two cloud-free acquisitions.

Satellite radargrammetry

The same principle of using two viewing geometries to help extract the geometry of the surface being viewed can be extended to active radar data (Toutin and Chénier, 2009). Radar data penetrate cloud cover and can even be used at night, so the technique compensates for two limitations of optical data. It can be used to generate DTMs at similar resolutions to optical satellites of areas under perennial cloud cover, where optical sensors cannot image the ground easily. The main constraint is that the algorithms are complex for most remote sensing practitioners to implement and rarely feature in off-the-shelf software packages. So, the technique is not widely applied.

Satellite radar interferometry InSAR

A more common way of exploiting satellite radar data to extract changes in digital terrain models is InSAR (Ferretti et al., 1999). This also exploits two images, but separated in time rather than space. The technique relies on measuring a change between acquisitions in the number of radar waves from a reflector on the ground to the sensor in orbit. The simplest analogy is the radar gun used by police to measure the speed of approaching vehicles. In practice, the two images are interfered and a series of fringes result, whose width is governed by the radar wavelength. These can be thought of as contours of ground motion. Because the technique measures only the change in ground position, an existing terrain model is used for the baseline digital terrain models, against which to measure the change.

This technique is powerful because it directly measures the change in ground position as a fraction of the radar's wavelength, irrespective of its spatial resolution. So, while the spatial resolution may be 25 m, the sensitivity to ground movement is typically on the order of a few cm, depending on the wavelength used. This matches closely the order of magnitude of many ground movements, and so the technique has been applied to a wide range of such phenomena, including seismic activity, volcanic uplift, swelling and shrinking clays, and groundwater and mineral extraction. It is even more powerful when combined with a geodetic surveying technique such as GPS. InSAR gives a wide area, synoptic view of motion restricted to specific time periods. GPS is a spatially discrete, sparse measurement but can be acquired almost constantly. It can calibrate InSAR and partially fill gaps between InSAR acquisitions.

InSAR, inevitably, has several significant weaknesses. Between acquisitions, surface characteristics other than the ground's position can change; if this change is large, the two images become incomparable (they are referred to as having low coherence) and the technique does not work. Changing vegetation is a particular challenge for short-wavelength, C-band radar, and so InSAR works best in arid countries and urban areas. Longer wavelengths such as L band are less affected by vegetation, but they are also less sensitive to ground motion. Changes occur in atmospheric properties, too, and these must be removed for the technique to work. Finally, properly calibrated measurements are still in the line of sight of the sensor; this matters more for radar data than optical data, because radars are acquired with a look angle of less than 90° to the horizontal. So, the result must be transformed into x, y, and z coordinates, either by using near-contemporary image pairs from both ascending and descending satellite passes with different geometries or by integrating InSAR with data like GPS.

Persistent Scatterer Interferometry PSInSAR

To get around the coherence limitation, a technique known as PSInSAR exploits strong radar reflectors such as building corners and metal objects that persist over time (i.e., are coherent) (Ferretti et al., 2000). In this technique,

several tens of images are acquired over a decade or more and processed together to extract these persistent scatterers (PS), which are then the focus for the interferometric processing. Rather than fringes over wide areas, this results in ground movement histories for each PS identified. These can then be gridded and interpolated to make an image, typically shown as mean annual velocity or viewed as graphs of ground movement against time for individual PS. Statistics dictate that the use of large numbers of images leads to enhanced resolution, on the order of 2 mm/year for C-band data. Constraints are that the measurement is again in the line of sight of the radar, atmospheric changes must be screened, orbital effects must be removed, and the technique cannot measure ground movements the size of multiple radar wavelengths, as this leads to aliasing. So, applications like mining with larger ground movements require caution. There is also no guarantee that PS will occur where investigators want them to. This can be overcome by placing metal corner reflectors or active transponders in the study area on the features of interest, such as reservoirs or bridges (Xia et al., 2002), to act as well-located persistent scatterers.

Airborne measurement techniques

All of the above techniques can be applied from an airborne platform. The biggest advantages that this brings are increased spatial resolution and the ability to target good weather windows. A particularly powerful technique is laser ranging, or lidar (Liu, 2008). The increased resolution that results from flying much closer to the target means that cm resolution lidar DTMs can be acquired and compared. Digital terrain models at tens of cm resolution are routinely generated from stereo aerial photography (Fabris and Pesci, 2005), and repeat airborne surveys can be used to measure changes in DTM in exactly the same way as for stereo optical satellite data. Radar interferometry can also be performed from an airborne platform to give, and compare, 1 m or better DTMs (Madsen et al., 1993). The natural geometric variability of flight lines flown through the lower atmosphere is greater than that for the orbits of satellites, and so this places a high premium on geocorrection and registration. But, at these high resolutions, the major constraints are vegetation and buildings, because the DTM measures the first surface a signal is returned from, such as the top of the tree canopy or the roofs of buildings. Corrections must be applied to remove these features and produce a bare earth model of the ground. Buildings are predictable enough to be removed cleanly, but trees are variable in height and much harder to deal with. Of the three techniques, lidar comes into its own here, because lidar with sufficiently high point densities that record the entire spectrum of returns ensures that at least some of the returns come from the ground, in between the trees. By always taking the last return, a bare earth model can be constructed even in fairly heavily forested terrain (Liu, 2008).

Ground-based measurement techniques

All of the above techniques can also be applied from a ground-based platform. The biggest advantages that this brings are further increases in spatial resolution and even greater flexibility to target weather windows. Such high-resolution surveys naturally cover smaller study areas and so are best targeted on ground motion of known extent on a site-specific basis. Terrestrial lidar has been used extensively to monitor slope stability (Jaboyedoff et al., 2012), with the further advantage that vertical surfaces like cliffs that are not seen in airborne or satellite datasets can be imaged easily. When combined with photography, the models produced can provide detailed and realistic views of the surfaces under investigation. This may allow more to be said about the causes of ground movement, rather than simply measuring the motion. Terrestrial radar interferometry is starting to be used in similar ways; its ability to penetrate clouds means that it has particular utility in volcano monitoring (Rödelberger et al., 2010). One drawback is that, even for restricted study areas, these ground-based sensors generate very large datasets that have a large processing, manipulation, and storage overhead.

Interpreting ground movement measured in remote sensing data

Remote sensing measures the movement of the ground at the surface. As we have seen, this can have many and even multiple causes. For example, an average ground movement map from PSInSAR over a typical city shows motion that has a variety of causes. There may be settlement of alluvium along major river valleys, subsidence above tunneling and subsidence due to groundwater abstraction. In addition there may be evidence of tectonic motion and almost certainly some swelling and shrinking clays. The remote sensing technique does not discriminate, but merely measures the ground movement in each case. The interpretation of the data requires interaction between the remote sensing expert and the geoscientist, if the causes of motion are to be fully understood. In fact, this interaction is needed from the outset, so the best technique can be selected to measure the particular type of movement that can be expected to be found in the study area. Study of the ground movement should also go beyond measurement of the motion. All of these remote sensing datasets tell us something about the geology exposed at the surface and can be interpreted in terms of faulting, folding, and lithology, for example. Such interpretation helps us understand the observed ground movement. In addition, it may be necessary to use other types of remote sensing that do not measure the ground motion at all, but do provide further insights into the geology, such as multi- or hyperspectral optical data for lithology and mineralogy. In addition, the remote sensing should be used in combination with other techniques that are not restricted to measuring surface phenomena, such as geophysics, in order to understand ground movement in the full three dimensions in which it occurs.

Summary and conclusions

Ground movements occur for a variety of reasons including natural geological processes such as landsliding and anthropogenic activities such as mining. They have a serious impact in terms of financial losses, injuries, and even deaths. Remote sensing can measure the movement of the ground in a variety of ways, all of which can be applied from space, aircraft, or in field surveys. These include photogrammetry with both optical and radar data, radar interferometry, and ranging techniques, such as lidar. The measurement of ground movement requires two acquisitions separated in time and can be extended into the monitoring of ground motion by repeat surveys.

Bibliography

- Bell, F. G., 1998. *Environmental Geology: Principles and Practice*. Oxford: Blackwell Science. 594p.
- Bell, F. G., 1999. *Geological Hazards. Their Assessment, Avoidance and Mitigation*. London: E & F N Spon. 648p.
- Bell, F. G., 2007. *Engineering Geology*, 2nd edn. Oxford: Butterworth-Heinemann. 581p.
- Bell, F. G., and Donnelly, L. J., 2006. *Mining and Its Impact of the Environment*. London: Taylor & Francis. 547p.
- Cheetham, R., 2008. Subsidence: a gradual catastrophe. *Catastrophe Risk*. December 2008.
- Cheney, R., Douglas, B., Agreen, R., Miller, L., Milbert, D., and Porter, D., 1986. The GEOSAT Altimeter Mission: a milestone in satellite oceanography. *Eos, Transactions American Geophysical Union*, **67**(48), 1354–1355, doi:10.1029/EO067i048p01354. <http://dx.doi.org/10.1029/EO067i048p01354>
- Fabris, M., and Pesci, A., 2005. Automated DEM extraction in digital aerial photogrammetry: precisions and validation for mass movement monitoring. *Annals of Geophysics*, **48**(6), 973–988, doi:10.4401/ag-3247. <http://www.annalsofgeophysics.eu/index.php/annals/article/view/3247>
- Ferretti, A., Prati, C., and Rocca, F., 1999. Multibaseline INSAR DEM reconstruction, the wavelet approach. *IEEE Transactions on Geoscience and Remote Sensing* **37**(2), 705–715, doi:10.1109/36.752187. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=752187&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D752187
- Ferretti, A., Prati, C., and Rocca, F., 2000. Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, **38**(5), 2202–2212, doi:10.1109/36.868878. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=868878&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D868878
- Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M.-H., Loye, A., Metzger, R., and Pedrazzini, A., 2012. Use of LIDAR in landslide investigations: a review. *Natural Hazards*, **61**(1), 5–28, doi:10.1007/s11069-010-9634-2. <http://link.springer.com/article/10.1007/s11069-010-9634-2>
- Johnson, R. B., and DeGraff, J. V., 1988. *Principles of Engineering Geology*. New York: Wiley. 497p.
- Keller, E. A., 1985. *Environmental Geology*, 4th edn. Columbus, OH: Charles E Merrill Publishing Company. 480p. (Note: 8th Edition 1999).
- Keller, E. A., Blodgett, R. H., and Clague, J. J., 2008. *Natural Hazards: Earth's Processes as Hazards, Disasters and Catastrophes*, Canadian edn. Toronto: Pearson. 448p.
- Liu, X., 2008. Airborne LiDAR for DEM generation: some critical issues. *Progress in Physical Geography*, **32**(1), 31–49, doi:10.1177/0309133308089496. <http://ppg.sagepub.com/content/32/1/31.abstract>

- Madsen, S. N., Zebker, H. A., and Martin, J., 1993. Topographic mapping using radar interferometry: processing techniques. *IEEE Transactions on Geoscience and Remote Sensing*, **31**(1), 246–256, doi:10.1109/36.210464. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=210464&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D210464
- McCall, G. J. H., Laming, D. J. C., and Scott, S. C. (eds.), 1992. *Geohazards Natural and Man-Made*. Chapman & Hall: London. 227p.
- Nuhfer, E. B., Proctor, R. J., Moser, P. H., et al., 1993. *The Citizens' Guide to Geologic Hazards*. Arvada, CO: The American Institute of Professional Geologists. 134p.
- Rödelberger, S., Läuter, G., Gerstenecker, C., and Becker, M., 2010. Monitoring of displacements with ground-based microwave interferometry: IBIS-S and IBIS-L. *Journal of Applied Geodesy*, **4**(1), 41–54, doi:10.1515/jag.2010.005, ISSN (Online) 1862-9024, ISSN (Print) 1862-9016. [http://www.degruyter.com/dg/viewarticle/j\\$002fjag.2010.4.issue-1\\$002fjag.2010.005\\$002fjag.2010.005.xml](http://www.degruyter.com/dg/viewarticle/j$002fjag.2010.4.issue-1$002fjag.2010.005$002fjag.2010.005.xml)
- Toutin, T., 2001. Elevation modelling from satellite visible and infrared (VIR) data. *International Journal of Remote Sensing*, **22**(6), 1097–1125, doi:10.1080/01431160117862. <http://www.tandfonline.com/doi/abs/10.1080/01431160117862#UfKhldwah0>
- Toutin, T., and Chenier, R., 2009. 3-D radargrammetric modeling of RADARSAT-2 ultrafine mode: preliminary results of the geometric calibration. *Geoscience and Remote Sensing Letters, IEEE*, **6**(2), 282, 286, doi:10.1109/LGRS.2008.2010563. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4768710&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxppls%2Fabs_all.jsp%3Farnumber%3D4768710
- Turner, A. K., and Schuster, R. L. (eds.), 1996. *Landslides: Investigation and Mitigation*. Washington, DC: National Research Council. Transport Research Board Special Publication, Vol. 247. 675p.
- Waltham, A. C., Bell, F. G., and Culshaw, M. G., 2005. *Sinkholes and Subsidence: Karst and Cavernous Rocks in Engineering Construction*. Berlin/Chichester: Springer/Praxis Publishing. 382p.
- Xia, Y., Kaufmann, H., and Guo, X., 2002. Differential SAR interferometry using corner reflectors. In *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02*. IEEE International, Vol. 2, 1243–1246, doi:10.1109/IGARSS.2002.1025902. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1025902&sortType%3Dasc_p_Sequence%26filter%3DAND%28p_IS_Number%3A22037%29

SURFACE RADIATIVE FLUXES

Rachel T. Pinker
Department of Atmospheric and Oceanic Science,
University of Maryland, College Park, MD, USA

Synonyms

Remote sensing of surface radiative fluxes; Shortwave and longwave surface radiative fluxes

Definition

Radiative fluxes. Electromagnetic radiation received from the sun (shortwave) (SW) or emitted from the atmosphere and/or the Earth surface (longwave) (LW).

Surface radiative fluxes. Electromagnetic radiation received at the surface of the Earth (ocean, land, and cryosphere) in the shortwave or longwave part of the spectrum.

Surface radiation balance. The sum of the incoming and outgoing shortwave and longwave radiative fluxes at the surface.

Remote sensing of radiative fluxes. Sensing of radiative fluxes by instruments away from the source of radiation such as satellites or aircraft.

Introduction

Components of surface radiative fluxes

Solar radiation (*insolation or shortwave radiation*) from the sun (0.3–4.0 μm) that reaches the Earth's surface (about 50 % of that emitted) is a major source of energy for heating our planet. At the mean distance of the Earth from the sun (1.50×10^8 km), the incident radiant flux density (*irradiance*) on a surface perpendicular to the solar beam is known as the *solar constant* quoted to be about $1,360.8 \pm 0.5 \text{ W m}^{-2}$ according to measurements from the Total Irradiance Monitor (TIM) on NASA's Solar Radiation and Climate Experiment (SORCE) (Kopp et al., 2005; Kopp and Lean, 2011). Solar radiation at any location on the Earth is determined by the *Earth–Sun distance* and the *solar zenith angle*. On its way from its source, it is partially absorbed and scattered by the atmosphere and clouds and partially absorbed by the surface. It arrives at the surface as *direct solar* and *diffuse sky radiation*; the sum is referred to as *global radiation*. The absorbed energy in the atmosphere reradiates back to the surface and to outer space in the form of longwave (*terrestrial*) radiation (4.0–100.0 μm), and similarly, part of the absorbed energy at the surface is reradiated back to the atmosphere and space. The Earth climate is a result of the balance maintained between the *solar radiation* absorbed by the Earth–atmosphere system (gain) and the *emission* of the *terrestrial* radiation back to space (loss). This balance is also referred to as *net radiation* (Q^*) or *radiation balance* composed of *net solar radiation* and *net longwave radiation* expressed commonly as

$$Q^* = \text{SW}\downarrow - \text{SW}\uparrow + \text{LW}\downarrow - \text{LW}\uparrow \quad (1)$$

where SW (0.3–4.0 μm) and LW (4.0–100.0 μm) are the shortwave and longwave components, respectively, and arrows represent the direction of the flux and Q^* is the surface net radiation (0.3–100.0 μm). The net shortwave radiation is a balance between the incoming shortwave and the reflected shortwave; the ratio of reflected shortwave to incoming shortwave radiation is also known as the *albedo* (A):

$$K^* = \text{SW}\downarrow(1 - A) \quad (2)$$

K^* is the surface *net* shortwave radiation (0.3–4.0 μm).

The net longwave radiation is a balance between the incoming longwave and outgoing longwave; the latter depends on *surface temperature* and *emissivity*:

$$L^* = (LW\downarrow - LW\uparrow) \quad (3)$$

L^* is net longwave radiation at the surface (4.0–100.0 μm), and $LW\uparrow$ can be determined as

$$LW\uparrow = \varepsilon\sigma T^4 \quad (4)$$

where ε is surface emissivity, σ is the Stefan–Boltzmann constant, and T is the surface temperature.

Absorption of solar radiation in the atmosphere is mostly from ozone, water vapor, carbon dioxide, oxygen, and clouds. Clouds play a major role in determining the net radiative balance via their optical properties and amount. The effect of clouds on the Earth's radiation balance is measured as the difference between the clear-sky and total-scene radiation results. This difference is defined as *cloud-radiative forcing*. *Cloud optical depth* is a general measure of the capacity of a cloud to control the amount of light that will reach the surface. Greater optical depth means greater blockage of the light and a larger cooling of the Earth–atmosphere system in the SW region of the spectrum.

Attempts to estimate the radiative components of the Earth–atmosphere system date as early as Budyko (1958). Annually and globally averaged climatological values as estimated by Kiehl and Trenberth (1997) are 342 W m^{-2} for incoming solar, 107 W m^{-2} for reflected solar (namely, about 30 % of the incoming solar energy is reflected back to space), and 235 W m^{-2} for outgoing longwave. These values are being continuously reevaluated based on improved satellite observations (e.g., Hansen et al., 2005; Trenberth et al., 2009; Ma and Pinker, 2012). For some time, it was believed that about 20 % is absorbed by atmospheric ozone, water vapor, carbon dioxide, oxygen, and clouds. Recent investigations based on field experiments and numerical computations indicate that the atmosphere absorbs more than 25 % of the solar radiation due to several factors, e.g., the neglect of absorption in the wavelength range beyond 2.8 μm (Cess et al., 1997; Valero et al., 2003). The effect of clouds on atmospheric absorption is small, about 3 W m^{-2} .

In reality, the energy balance distribution over the globe is not uniform, with a net energy gain in lower latitudes and a loss at higher latitudes. One major objective of remote sensing from satellites is to obtain the best possible information on the radiation balance of the Earth–atmosphere system (Trenberth et al., 2009).

Need for information on radiative fluxes

Accurate information on the various components of the radiative fluxes at global scale is needed for a wide range of climate research applications such as modeling the hydrological cycle, representing interactions and feedbacks between the atmosphere and the surface, estimating

global terrestrial and oceanic net primary productivity, and validating climate and numerical weather prediction models. Each component has significance for a different type of application. For example, net all-wave radiation (0.3–100.0 μm) is needed for hydrological modeling, photosynthetically active radiation (PAR) (0.4–0.7 μm) for modeling net primary productivity and CO_2 budgets, and surface albedo for climate change studies. Satellite observations provide a systematic source of information at global scale needed for inferring components of the radiation balance. Recent advances in satellite technology and development of inference schemes have been conducive for obtaining information on the various components of the radiation balance. Reviewed will be basic principles of methodologies used to derive the Surface Radiation Budget by methods of remote sensing, feasibility to derive such fluxes and linkage to international programs, evaluation efforts, and future prospects.

Type of satellites used for estimating radiative fluxes

Information on radiative fluxes is being derived from dedicated satellites as well as from operational weather satellites. Both observational systems are composed of two types of satellites: geostationary and polar orbiting. The U. S. Geostationary Operational Environmental Satellites are known as GOES, while the Polar-Orbiting Environmental Satellites are referred to as POES. GOES satellites circle the Earth in a geosynchronous orbit, namely, they orbit the equatorial plane of the Earth at about 35,800 km above the Earth at a speed matching the Earth's rotation. The United States normally operates two geostationary satellites in orbit over the equator, one is positioned at 75° W while the other at 135° W longitudes. Currently, the United States is operating GOES-15 (or GOES-West) and GOES-13 (or GOES-East). Complementing the geostationary satellites are two polar-orbiting ones constantly circling the Earth in an almost north-south orbit, passing close to both poles. The orbits are circular, with an altitude between 830 and 870 km, and are sun synchronous. Environmental satellites are also operated by other nations and international organizations such as METEOSAT Second Generation (MSG) operated by EUMETSAT; LANDSAT, a high-resolution Earth observation satellite managed by NASA and the USGS; METOP, the European polar orbiter; and MTSAT, a Japanese geostationary satellite.

Radiative fluxes by remote sensing: feasibility

Satellites can provide information on surface radiative fluxes at various temporal and spatial scales, and attempts to do so have been ongoing for about 40 years. Steering has been provided by the World Climate Research Programme (WCRP)/World Meteorological Organization (WMO) and by several national and international projects, such as the Global Energy and Water Budget Experiment (GEWEX), the International Biosphere/Geosphere Programme (IGBP), and the GEWEX Hydroclimatology Panel (GHP) which coordinates scientific issues related

to the development and implementation of Regional Hydroclimate Projects. Specifications of requirements on information on radiative fluxes are provided in the Inter-governmental Panel on Climate Change (IPCC, 2007). In the United States, the *Interagency Committee on Earth and Environmental Sciences* has identified clouds and the hydrological cycle to be of highest scientific priority for global change research. Under this framework efforts are in progress to derive clouds and radiative fluxes from satellite observations. These include the *Surface Radiation Budget* (SRB) activity, the *International Satellite Cloud Climatology Project* (ISCCP) (Rossow and Schiffer, 1999), the *Clouds and the Earth's Radiant Energy System* (CERES) (Wielicki et al., 1996), the *Geostationary Earth Radiation Budget* (GERB) Experiment on the EUMETSAT METEOSAT Second Generation satellites (Harries et al., 2005), and the Moderate Resolution Imaging Spectroradiometer (MODIS) (Platnick et al., 2003; King et al., 2003). The instrument packages on these satellites are known as *imagers* (*sounders* are used for inferring the vertical structure of the atmosphere) and usually have about 2–36 channels in the spectral interval of 0.3–12.0 μm that are relevant for estimating both for shortwave and longwave fluxes.

Examples of channels on selected Geostationary Operational Environmental Satellites (GOES) are presented in Table 1. The channels of the Advanced Very High Resolution Radiometer (AVHRR) on polar-orbiting satellites are presented in Table 2. METEOSAT-5 and METEOSAT-7 channels are centered at 0.75, 6.4, and 11.5 μm , while the newer SEVIRI instrument on METEOSAT-8 has 12 channels as given in Table 3. The objective of NASA's *Earth Observing System* (EOS) Program is to study the Earth from space using a multiple-instrument, multiple-satellite approach. The *Clouds and the Earth's Radiant Energy System* (CERES) is one of the highest priority scientific satellite instruments developed for EOS (Wielicki et al., 1996; Loeb et al., 2009). CERES products include both solar-reflected and Earth-emitted radiation. The CERES instrument builds upon previous similar missions such as the *Earth Radiation Budget Experiment* (ERBE). CERES instruments were launched aboard the *Tropical Rainfall Measuring Mission* (TRMM) in November 1997 and on the EOS Terra satellite in December 1999. Two additional instruments fly on the EOS Aqua spacecraft since 2002. Each CERES instrument has three channels: a shortwave to measure reflected sunlight, a longwave to measure emitted thermal radiation in the 8.0–12.0 μm “window” region, and a total channel to measure all wavelengths of radiation (0.30–100.0 μm). The CERES instrument is managed by the NASA Langley Research Center in Hampton, Virginia.

The *Geostationary Earth Radiation Budget* (GERB) instrument is similar to CERES for the EUMETSAT's METEOSAT Second Generation (MSG) satellites intended to make accurate measurements of the Earth Radiation Budget from geostationary orbit. It was

Surface Radiative Fluxes, Table 1 Summary of spectral bands on several GOES imagers

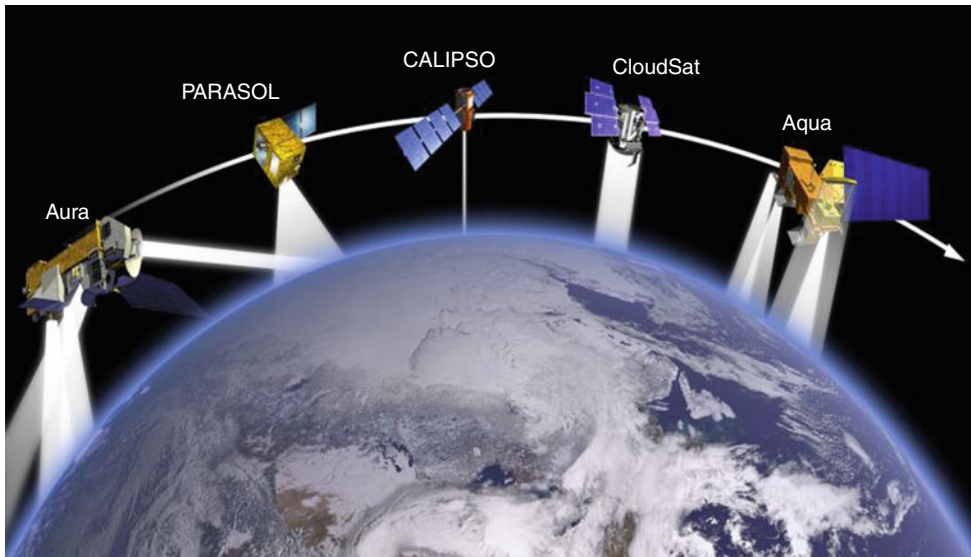
Band	Wavelength range (μm)	Central wavelength (μm)	Meteorological objectives
1	0.55–0.75	0.65	Daytime cloud cover and surface features
2	3.8–4.0	3.9	Low cloud/fog and fire detection
3	6.5–7.0	6.75 (GOES-8/11)	Upper-level water vapor
4	5.8–7.3	6.48 (GOES-12)	Surface or cloud-top temperature
5	10.2–11.2	10.7	Surface/cloud-top temperature and low-level water vapor
6	11.5–12.5	12.0 (GOES-8/11)	CO ₂ band: cloud detection

Surface Radiative Fluxes, Table 2 Summary of spectral bandwidths (μm) of several AVHRR sensors

Channel	TIROS-N	NOAA-6, -8, -10	NOAA-7, -9, -11, -12, -14	NOAA-13
1	0.55–0.90	0.58–0.68	0.58–0.68	0.58–0.68
2	0.725–1.10	0.725–1.10	0.725–1.10	0.725–1.0
3	3.55–3.93	3.55–3.93	3.55–3.93	3.55–3.93
4	10.5–11.5	10.5–11.5	10.3–11.3	10.3–11.3
5	Same as channel 4	Same as channel 4	11.5–12.5	11.4–12.4

Surface Radiative Fluxes, Table 3 Summary of the bands on the SEVIRI

Band	Wavelength (μm)	Central wavelength (μm)
1	0.56–0.71	0.635
2	0.74–0.88	0.81
3	1.50–1.78	1.64
4	3.48–4.36	3.90
5	5.35–7.15	6.25
6	6.85–7.85	7.35
7	8.30–9.10	8.70
8	9.38–9.94	9.66
9	9.80–11.80	10.80
10	11.00–13.00	12.00
11	12.40–14.40	13.40
12	HRV: broadband (about 0.4–1.1)	



Surface Radiative Fluxes, Figure 1 The Afternoon Constellation ("A-Train") consisting of five US and international satellites (Credit: NASA).

launched in January 2004 on the first METEOSAT Second Generation satellite (MSG-1), renamed to METEOSAT-8. It complements the MSG SEVIRI instrument (Schmetz et al., 2002). The development, integration, and flight of GERB on MSG are supported by EUMETSAT and the European Space Agency (ESA). GERB has capabilities to measure the shortwave and longwave radiation from the Earth every 15 min and carries a *black body* for thermal calibration and a *solar diffuser* for shortwave calibration.

A recent addition to the suite of satellites relevant for *Earth Radiation Budgets* is the CloudSat mission (Stephens et al., 2002). CloudSat is a NASA Earth System Science Pathfinder mission to provide observations necessary to advance our understanding of cloud abundance, distribution, structure, and radiative properties. CloudSat is complimented with the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) (launched in April 2006) (Winker et al., 2012) and subsequently joined with several satellites already in orbit to form a constellation of satellites known as the *Afternoon Constellation* or *A-Train* (Figure 1). It consists of five satellites flying in close proximity: Aqua, CloudSat, CALIPSO, PARASOL, and AURA. The orbits are sun synchronous and provide near-simultaneous and colocated observations conducive to improved information on the *Earth Radiation Budget and Atmospheric Chemistry* (primarily from Aura).

The *Advanced Baseline Imager* (ABI) is being developed as the future *imager* on the Geostationary Operational Environmental Satellite (GOES) series, slated to be launched approximately in 2015 with GOES-R (Schmit et al., 2005). The ABI will begin a new era in US environmental remote sensing with more spectral

bands, faster imaging, and higher spatial resolution than the current imagers (Schmidt et al., 2005). ABI spatial resolution will be nominally 2 km for the infrared bands and 0.5 km for the 0.64 μm visible band. ABI has 16 spectral bands; 5 are similar to the 0.6, 4.0, 11.0, and 12.0 μm windows and the 6.5 μm water vapor band on several *GOES* imagers (e.g., 8/-9/-10/-11) (Menzel and Purdom, 1994) and another is similar to the 13.3 μm on the *GOES-12-13-14-15* imagers. Indeed, multiple satellites are needed to provide adequate temporal sampling since clouds and radiative fluxes vary throughout the day.

Physical principles of inference schemes

SW radiation

An important physical principle for deriving *SW* budgets from satellite observations is the close linear coupling between *SW* (0.3–4.0 μm) reflected radiance at the top of the atmosphere (*albedo*) and the surface irradiance (Schmetz, 1989). Cloud absorption (*transmittance*) and albedo are linearly related since atmospheric constituents do not emit radiation at solar wavelengths. There is a dependence on solar zenith angle, gaseous and aerosol absorption and scattering, surface reflectivity, and clouds. A plausible scenario of a retrieval methodology of shortwave radiative fluxes both at the surface and at the top of the atmosphere (TOA) can be summarized as follows. TOA downward flux (F_{td}) can be calculated from the extraterrestrial solar spectrum, accounting for variation in Sun–Earth distance and position of the sun in the sky relative to local vertical (*solar zenith angle*). Downward flux at the surface (F_{sd}) can be obtained by determining what fraction of F_{td} reaches the surface as radiation is

transferred through the atmosphere. This fraction is referred to as flux *transmittance* (T). It can be determined from a combination of the reflected part at the radiative flux at the TOA that the satellite observes and from modeling the radiative effects of atmospheric constituents (such as ozone, water vapor, molecular scattering, and optical properties of clouds and aerosols). It also depends on the length of the path the radiation travels through the atmosphere (determined by the solar zenith angle), and to a lesser degree, on the *albedo* of the surface. Once T is known, surface downward flux is obtained as $F_{sd} = T F_{td}$. Once F_{sd} is known, upward flux at the surface (F_{su}) is calculated as $F_{su} = A_s F_{sd}$, where A_s is surface albedo. Similarly, flux reflected to space by the Earth-atmosphere system (TOA upward flux, F_{tu}) can be obtained from the product of F_{td} and the TOA albedo (A_t), namely, $F_{tu} = A_t F_{td}$. For example, the Clouds and the Earth's Radiant Energy System (CERES) instrument onboard several Earth Observing System (EOS) satellites provides a direct measurement of TOA broadband (A_t).

LW radiation

Downwelling longwave radiation $LW\downarrow$ at the surface originates from radiatively active gases in the atmosphere and depends on the vertical profiles of temperature, gaseous absorbers, and clouds. Gaseous emission is present in specific wavebands in the range 0.3–30 μm , whereas emission from clouds corresponds to black body radiation at the temperature of the cloud base. The intensity of the thermal emission from a cloud varies with its temperature and the optical thickness of the cloud.

Retrieval of $LW\downarrow$ is difficult since only *atmospheric window* radiances at the top of the atmosphere give information on the near-surface radiation field. For the remainder of the longwave spectrum, the radiation regimes at the top of the atmosphere and at the surface are decoupled; more than 80 % of the clear-sky longwave flux reaching the surface is emitted within the lowest 500 m of the atmosphere. Cloud contributions are mainly from the atmospheric window region (8.0–13.0 μm), and the relevant cloud parameters are cloud base height location and temperature, *emittance*, and cloud amount. Relative importance of cloud contribution decreases with moister atmospheres since the transparency of the window decreases due to water vapor *continuum* absorption. From cloudless skies, more than half the longwave flux received at the ground comes from gases in the lowest 100 m and roughly 90 % from the lowest kilometer. Consequently, $LW\downarrow$ is often estimated as a function of a bulk atmospheric temperature (in Kelvin) approximated by air temperature at the ground and an estimated broadband atmospheric emissivity (Prata, 1996; Dille and O'Brien, 1998). It is often assumed that the longwave emissivity of most natural surfaces is unity, so that $L\uparrow$ can be estimated from knowledge of the surface temperature. Longwave techniques could be characterized as being of distinct types:

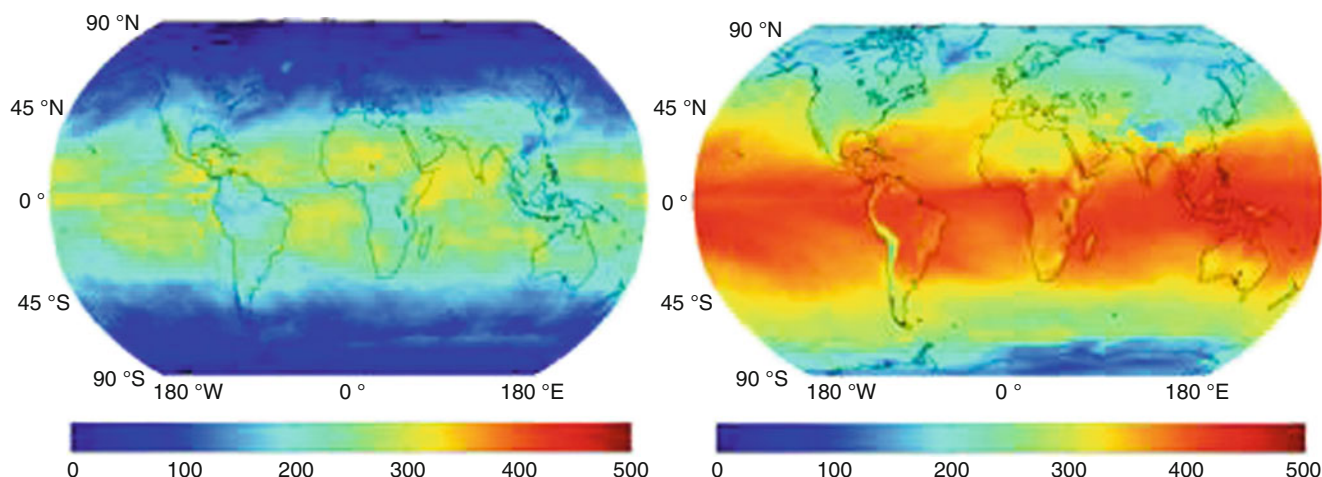
flux calculation techniques, which use radiative transfer models to calculate the $LW\downarrow$ directly from retrieved profiles of temperature and water vapor and estimates of cloud fraction and cloud base altitude (e.g., Breon et al., 2002; Gupta et al., 2010), flux inference techniques which use direct inference from *TOVS* or *HIRS* radiance observations, and statistical inference schemes based on radiance–flux relationship developed with detailed radiative transfer models and large number of observed atmospheric soundings (for clear-sky and overcast conditions) with some built in cloud height and base information. It is often assumed that the longwave emissivity of most natural surfaces is unity, so that $LW\uparrow$ can be estimated from knowledge of the surface temperature (Prata, 1996).

Methodologies to derive radiative fluxes

Inference schemes to derive radiative fluxes from satellite observations have been implemented at a wide range of complexities, both in the SW and LW part of the spectrum. Review type of summaries can be found in Schmetz (1989, 1991), Pinker et al. (1995), and Whitlock et al. (1995); methods to derive *PAR* are described in Frouin and Pinker (1995); recent approaches for *LW* methods can be found in Zhou et al. (2007), Nussbaumer and Pinker (2011, 2012), and Niemela et al. (2001).

Methods to derive SW fluxes have been implemented on different spatial and temporal scales. Used are observations from geostationary satellites (METEOSAT, GOES, GMS) and from polar-orbiting satellites using instruments such as AVHRR, MODIS, and CERES (Stuhlman et al., 1990; Pinker and Laszlo, 1992; Gupta et al., 1999; Chou et al., 1995; Brisson et al., 1994; Li et al., 1993; Zhang et al., 1995; Ceballos et al., 2004; Hollmann et al., 1999; Zhang et al., 2004; Charlock et al., 2006; Vardavas and Taylor, 2007; Wang and Pinker, 2009; Ma and Pinker, 2012). For obtaining estimates at global scale at three hourly intervals, the International Satellite Cloud Climatology Project (ISCCP) (Rossow and Schiffer, 1991) has merged information from several satellites that has been widely used and implemented with independent algorithms. Initial work on the use of these observations started at the NASA Goddard Institute for Space Studies (GISS) (Rossow and Lacis, 1990). The validation results for the GISS ISCCP surface fluxes show that it has an uncertainty of 10–15 W m^{-2} at monthly time scales (Zhang et al., 2006, 2007). The ISCCP data in their various versions have been widely used by many groups (Stackhouse et al., 2002; Ma and Pinker, 2012) and serve as the basis for the WCRP-sponsored *Radiative Flux Assessment* intercomparison activity under which various satellite products (global and regional) are being evaluated.

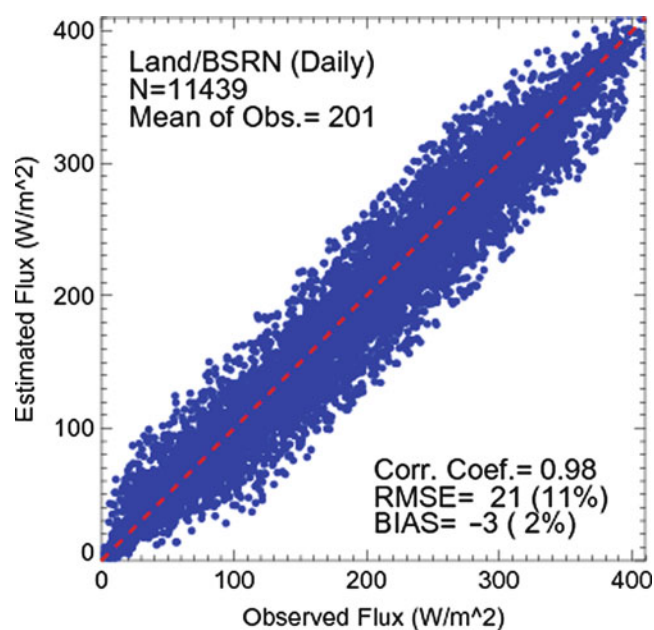
Many of the current geostationary satellites are limited in their capability to accurately detect cloud or aerosol optical properties that are important elements of the



Surface Radiative Fluxes, Figure 2 Left: Monthly averaged surface LW radiation for March 2001 as derived from MODIS observations (Nussbaumer and Pinker, 2012); Right: monthly averaged surface SW radiation for March 2001 as derived from MODIS observations (Wang and Pinker, 2009).

radiation budget. Polar-orbiting satellites tend to have higher spatial resolution than the geostationary satellites, as well as more spectrally resolving bands. The MODIS instrument onboard the Terra and Aqua satellites, a state-of-the-art sensor with 36 spectral bands and onboard calibration of both solar and infrared bands, has the capability to measure atmospheric and surface properties with higher accuracy and consistency than previous Earth observation imagers. Moreover, geostationary satellites cover only part of the globe up to about 50° north and south. Polar-orbiting satellites provide coverage twice a day over the entire globe, including the high latitudes. A configuration of satellites, both geostationary and polar orbiting, is needed to cover the entire globe with observations that depict the diurnal cycle. In Figure 2 shown are mean surface SW radiative fluxes (W m^{-2}) averaged for March 2001 (Wang and Pinker, 2008) and the mean LW $\downarrow$ fluxes for the same month (Nussbaumer and Pinker, 2012) as obtained from the MODIS instrument on Terra and Aqua.

Real-time implementation of inference schemes of shortwave radiative fluxes is now a reality. Flux estimates are produced at hourly time scale by the National Oceanic and Atmospheric Administration (NOAA)/National Environmental Satellite, Data, and Information Service (NESDIS), using observations from GOES (Pinker et al., 2003). Similar efforts are ongoing in other countries such as at the *Nowcasting Satellite Application Facility* (NWC SAF) in Europe. SAFs are dedicated centers for processing satellite data. They have been developed by a consortium of EUMETSAT Member States and Cooperating States trying to benefit from METEOSAT Second Generation (MSG) and Polar Platform Satellites.



Surface Radiative Fluxes, Figure 3 Evaluation of daily mean downwelling SW flux estimated by UMD/SRB_MODIS model (Wang and Pinker, 2009) against BSRN measurements over land (January 2003–December 2005). Cases eliminated: 1.6 %.

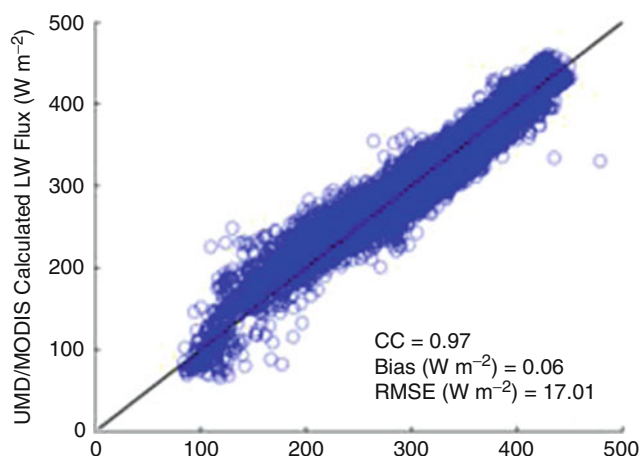
Evaluation of satellite estimates using surface measurements

Evaluation of surface fluxes depends on the quality of the surface measurements (Shi and Long, 2002; Philipona et al., 2001). Historic surface SW measurements were

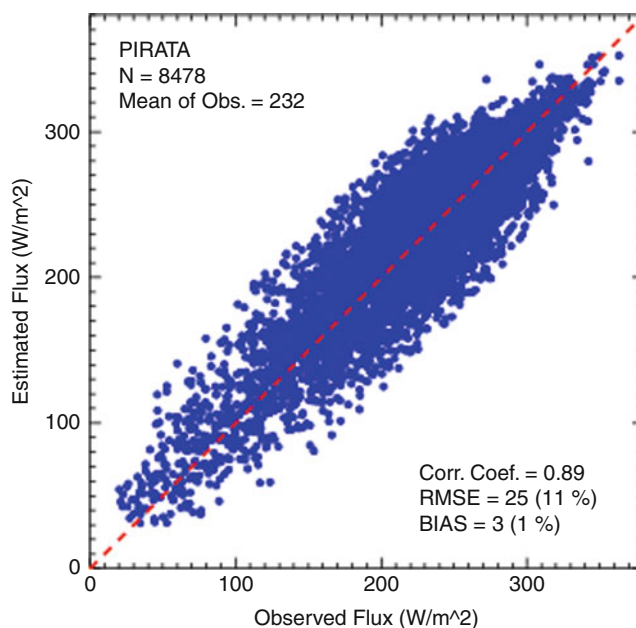
collected under the auspices of the WCRP at the World Radiation Data Centre (WRDC), A.I. Voeikov Main Geophysical Observatory, St. Petersburg, Russia. This database includes sites from all continents and contains measurements in the form of daily and monthly averages. The Swiss Federal Institute of Technology provides the Global Energy Balance Archive (GEBA) (Ohmura and Gilgen, 1993) which includes data from the WRDC. Globally distributed surface measurements over the last two decades are improving with the advent of the GEWEX Baseline Surface Radiation Network (BSRN) (Ohmura et al., 1998). BSRN is dedicated to providing long-term high-quality measurements at locations around the globe critical for satellite validation. Regional-scale networks of surface measurements are also utilized in validation (Augustine et al., 2000; Michalsky et al., 1999) as well as data from special campaigns such as the Clouds and the Earth's Radiant Energy System (CERES) program/Atmospheric Radiation Measurement (ARM) Validation Experiment (CAVE) (Rutan et al., 2001). Examples of evaluation of SW and LW results from MODIS observations against the BSRN network are shown in Figures 3 and 4. Radiative fluxes are also measured over the oceans. For instance, in situ measurements from the Pilot Research Moored Array in the Tropical Atlantic (PIRATA) moorings in the tropical Atlantic (Bourle's et al., 2008) and the Tropical Atmosphere Ocean/Triangle Trans-Ocean Buoy Network (TAO/TRITON) moorings in the tropical Pacific Ocean (McPhaden et al., 1998) are frequently used for comparisons with the satellite estimates (Figure 5). Surface radiative fluxes are needed at accuracies suitable to predict transient climate variations and long-term climate trends. The *WCRP GEWEX Surface Radiation Budget (SRB)* project is specially tasked to produce, validate, and assess long-term surface and atmospheric radiative budgets at global scale. Intercomparison between satellite products with surface sites is part of the GEWEX SRB (Stackhouse et al., 2002) and CERES-SARB programs (Charlock et al., 2006). Reported is agreement with ground observations within 10 W m^{-2} on a monthly time scale. At hourly time scales the reported differences are in the range of -50 to $+50 \text{ W m}^{-2}$. More comprehensive evaluation results are reported on in Trenberth et al. (2009), Ma and Pinker (2012), and Nussbaumer and Pinker (2012).

Future prospects

A summary of current and future satellite observations of relevance for SRB research is presented in Wielicki et al. (1996). It is believed that SRB activities could benefit from better information on the vertical structure of clouds and meteorology of the lower layer; information on cloud types; improved calibration and stability of all instruments used; better understanding of validation limitations; improvement in inference techniques; and information on aerosol optical properties, better characterization of



Surface Radiative Fluxes, Figure 4 Evaluation of daily mean surface LW flux estimated by UMD/SRB_MODIS_LW at 1° spatial resolution against 18 BSRN stations for 2007.



Surface Radiative Fluxes, Figure 5 Evaluation of daily mean downwelling SW estimated by UMD/SRB_MODIS model (Wang and Pinker, 2009) (2003–2005) against PIRATA buoy observations. Cases eliminated: 1.1 %.

land surface properties, accurate information on snow distribution, and distinction of partially cloud-filled satellite fields of view and the 3-D nature of clouds. In the short term, the SRB inference methods will benefit from activities like the CloudSat, the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations, and the entire constellation of *A-Train* satellites.

Acknowledgments

We thank NASA for the support of our work under grant NNX08AN40A from the Science Mission Directorate-Division of Earth Science and grant NNX13AC12G from the Energy and Water Cycle Study (NEWS) program. The work also benefited from the support under NOAA grant NA09NES4400006, the Cooperative Institute for Climate and Satellites (CICS) at the University of Maryland/ESSIC.

Bibliography

- Augustine, J. A., DeLuisi, J. J., and Long, C. N., 2000. SURFRAD—A national surface radiation budget network for atmospheric research. *Bulletin of the American Meteorological Society*, **81**, 2341–2357.
- Bourle's, B., et al., 2008. The Pirata Program: history, accomplishments, and future directions. *Bulletin of the American Meteorological Society*, **89**, 1111–1125.
- Breon, F. M., Buriez, J. C., Couvert, P., et al., 2002. Earth's atmosphere, ocean and surface studies, Book series: *Advances in Space*, **30**(11), 2383–2386. Article Number: PII S0273-1177(02)00078-9, doi:10.1016/S0273-1177(02)80282-4.
- Brisson, A., Le Borgne, P., Marsouin, A., and Moreau, T., 1994. Surface irradiance calculated from Meteosat sensor data during SOFIA-ASTEX. *International Journal of Remote Sensing*, **15**, 197–203.
- Budyko, M. I., 1958. The heat balance of the earth's surface. Washington, DC: Dept. of Commerce, Weather Bureau, 1958 – Earth temperature, pp. 259.
- Ceballos, J. C., Bottino, M. J., and de Souza, J. M., 2004. A simplified physical model for assessing solar radiation over Brazil using GOES 8 visible imagery. *Journal of Geophysical Research: Atmospheres*, **109**(D2), Art. No. D02211.
- Cess, R. D., Zhang, M. H., Potter, G. L., Alekseev, V., Barker, H. W., Bony, S., Colman, R. A., Dazlich, D. A., DelGenio, A. D., Deque, M., Dix, M. R., Dymnikov, V., Esch, M., Fowler, L. D., Fraser, J. R., Galin, V., Gates, W. L., Hack, J. J., Ingram, W. J., Kiehl, J. T., Kim, Y., LeTreut, H., Liang, X. Z., McAvaney, B. J., Meleshko, V. P., Morcrette, J. J., Randall, D. A., Roeckner, E., Schlesinger, M. E., Sporyshev, P. V., Taylor, K. E., Timbal, B., Volodin, E. M., Wang, W., Wang, W. C., and Wetherald, R. T., 1997. Comparison of the seasonal change in cloud-radiative forcing from atmospheric general circulation models and satellite observations. *Journal of Geophysical Research*, **102**, 16593–16603.
- Charlock, T. P., Rose, F. G., Rutan, D. A., Jin, Z., and Kato, S., 2006. The global surface and atmospheric radiation budget: An assessment of accuracy with 5 years of calculations and observations. In *Proceedings 12th Conference on Atmos. Radiation*, Madison, WI, July 10–14.
- Chou, M. D., Ridgway, W. L., and Yan, M. M. H., 1995. Parameterizations for water vapor IR radiative transfer in both the middle and lower atmospheres. *Journal of the Atmospheric Sciences*, **52**(8), 1159–1167, doi:10.1175/1520.
- Dilley, A. C., and O'Brien, D. M., 1998. Estimating downward clear sky long-wave irradiance at the surface from screen temperature and precipitable water. *Quarterly Journal of the Royal Meteorological Society*, **124**, 1391–1401.
- Frouin, R., and Pinker, R. T., 1995. Estimating Photosynthetically Active Radiation (PAR) at the Earth's surface from satellite observations. *Remote Sensing of Environment*, **51**.
- Gupta, S. K., Ritchey, N. A., Wilber, A. C., Whitlock, C. H., Gibson, G. G., and Stackhouse, P. W. Jr., 1999. A climatology of surface radiation budget derived from satellite data. *Journal of Climate*, **12**(8), 2691–2710, Part 2.
- Gupta, S. K., Kratz, D. P., Stackhouse, P. W., Jr., Wilber, A. C., Zhang, T., and Sothcott, V. E., 2010. Improvement of surface longwave flux algorithms used in CERES processing. *Journal of Applied Meteorology and Climatology*, **49**(7), 1579–1589, doi:10.1175/2010JAMC2463.1.
- Hansen, J. E., et al., 2005. Earth's energy imbalance: confirmation and implications. *Science*, **308**, 1431–1435, doi:10.1126/science.1110252.
- Harries, J. E., Russell, J. E., and Hanafin, J. A., et al., 2005. The geostationary Earth Radiation Budget Project. *Bulletin of the American Meteorological Society*, **86**(7), doi:10.1175/BAMS-86-7-945.
- Hollmann, R., Feng, J., Leighton, H. G., Mueller, J., and Stuhlmann, R., 1999. ScaRaB as a valuable tool for BALTEx and MAGS: Satellite applications for energy budgets and the hydrological cycle. *Advances in Space Research*, **4**(7), 955–958.
- Intergovernmental Panel on Climate Change (IPCC), 2007. Climate change 2007: The physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change, edited by Solomon, S. et al. Cambridge, UK: Cambridge University Press.
- Kiehl, J. T., and Trenberth, K. E., 1997. Earth's annual global mean energy budget. *Bulletin of the American Meteorological Society*, **78**, 197–208, doi:10.1175/1520.
- King, M. D., et al., 2003. Cloud and aerosol properties, perceptible water, and profiles of temperature and humidity from MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 442–458, doi:10.1109/TGRS.2002.808226.
- Kopp, G., and Lawrence, G., 2005. The Total Irradiance Monitor (TIM): instrument design. *Solar Physics*, **230**, 1–2, doi:10.1007/s11207-005-7446-4.
- Kopp, G., and Lean, J. L., 2011. A new, lower value of total solar irradiance: evidence and climate significance. *Geophysical Research Letters*, **38**, L01706, doi:10.1029/2010GL045777.
- Kopp, G., Lawrence, G., and Rottman, G., 2005. The Total Irradiance Monitor (TIM): science Results, *Solar Physics*, **230**, 129–140.
- Li, Z., Leighton, H. G., Masuda, K., and Takashima, T., 1993. Estimation of SW flux absorbed at the surface from TOA reflected flux. *Journal of Climate*, **6**, 317–330.
- List, R. J. (ed.), 1966. *Smithsonian Meteorological Tables*, 6th edn. Washington, DC: Smithsonian Institution.
- Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., Manalo-Smith, N., and Wong, T., 2009. Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *Journal of Climate*, **22**, 748–766, doi:10.1175/2008JCLI2637.1.
- Ma, Y., and Pinker, R. T., 2012. Modeling shortwave radiative fluxes from satellites. *Journal of Geophysical Research*, **117**, D23202, 1–19, doi:10.1029/2012JD018332.
- McPhaden, M. J., et al., 1998. The Tropical Ocean-Global Atmosphere observing system: a decade of progress. *Journal of Geophysical Research*, **103**, 14,169–14,240.
- Menzel, W. P., and Purdom, J. F., 1994. Introducing GOES-I: the first of a new generation of geostationary operational environmental satellites. *Bulletin of the American Meteorological Society*, **75**, 757–782.
- Michalsky, J., Dutton, E., Rubes, M., et al., 1999. Optimal measurement of surface shortwave irradiance using current instrumentation. *Journal of Atmospheric and Oceanic Technology*, **16**(1), 55–69, doi:10.1175/1520-0426.
- Niemela, S., Raisanen, P., and Savijarvi, H., 2001. Comparison of surface radiative flux parameterizations. Part I: longwave radiation. *Atmospheric Research*, **58**, 1–18.
- Nussbaumer, E. A., and Pinker, R. T., 2011. Estimating surface long-wave radiative fluxes at global scale. *Quarterly Journal of Royal Meteorological Society*, **137**, October 2011 A.

- Nussbaumer, E. A., and Pinker, R. T., 2012. Estimating surface longwave radiative fluxes from satellites utilizing artificial neural networks. *Journal of Geophysical Research*, **117**, D07209, doi:10.1029/2011JD017141.
- Ohmura, A., and Gilgen, H., 1993. Re-evaluation of the global energy balance. *Geophysical Monograph*, **75**, IUGG, Vol. 15, pp 93–110.
- Ohmura, A., Dutton, E. G., Forgan, B., Frohlich, C., Gilgen, H., Hegner, H., Heimo, A., Konig-Langlo, G., McArthur, B., Muller, G., Philipona, R., Pinker, R. T., Whitlock, C. H., Dehne, K., and Wild, M., 1998. Baseline surface radiation network (BSRN/WCRP): new precision radiometry for climate research. *Bulletin of the American Meteorological Society*, **79**, 2115–2136.
- Philipona, R., et al., 2001. Atmospheric longwave irradiance uncertainty: pyrgeometers compared to an absolute sky-scanning radiometer, atmospheric emitted radiance interferometer, and radiative transfer model calculations. *Journal of Geophysical Research*, **106**, 28, 129–28, 141, doi:10.1029/2000JD000196.
- Pinker, R. T., and Laszlo, I., 1992. Modeling surface solar irradiance for satellite applications on a global scale. *Journal of Applied Meteorology*, **31**, 194–211.
- Pinker, R. T., Laszlo, I., Whitlock, C. H., and Charlock, T. P., 1995. Radiative flux opens new window on climate research. *EOS*, **76**(15), 145–160.
- Pinker, R. T., Tarpley, J. D., Laszlo, I., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Lohmann, D., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., and Higgins, R. W., 2003. Surface radiation budgets in support of the GEWEX continental scale international project (GCIP) and the GEWEX Americas prediction project (GAPP), including the North American land data assimilation system (NLDAS). *Journal of Geophysical Research-Atmospheres*, **108**(D22), Art. No. 8844 19 Nov 2003.
- Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riedi, J. C., and Frey, R. A., 2003. The MODIS cloud products: algorithms and examples from Terra. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 459–473.
- Prata, A. J., 1996. A new long-wave formula for estimating downward clear-sky radiation at the surface. *Quarterly Journal of the Royal Meteorological Society*, **122**, 1127–1151.
- Rossow, W. B., and Lacis, A. A., 1990. Global, seasonal cloud variation from satellite radiance measurements, 2, Cloud properties and radiative effects. *Journal of Climate*, **3**, 1204–1253.
- Rossow, W. B., and Schiffer, R. A., 1991. ISCCP cloud data products. *Bulletin of the American Meteorological Society*, **72**(1), 2–20.
- Rossow, W. B., and Schiffer, R. A., 1999. Advances in understanding clouds from ISCCP. *Bulletin of the American Meteorological Society*, **80**, 2261–2287, doi:http://dx.doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2.
- Rutan, D. A., Rose, F. G., Smith, N. M., and Charlock, T. P., 2001. Validation data set for CERES surface and atmospheric radiation budget (SARB), *WCRP/GEWEX Newsletter*, **11**(1), 11–12.
- Schmetz, J., 1989. Towards a surface radiation climatology: retrieval of downward irradiance from satellites. *Atmospheric Research*, **23**, 287–321.
- Schmetz, J., 1991. Retrieval of surface radiation fluxes from satellite data. *Dynamics of the Atmospheres and Oceans*, **16**(1–2), 61–72.
- Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A., 2002. An introduction to Meteosat Second Generation (MSG). *Bulletin of the American Meteorological Society*, **83**, 977–992.
- Shi, Y., and Long, C. N., 2002. Techniques and methods used to determine the best estimate of radiation fluxes at SGP Central Facility. Paper presented at *12th ARM Science Team Meeting Proceedings*, St. Petersburg, FL, 8–12 Apr 2002.
- Shmit, T. J., Gunshor, M. M., Menzel, W. P., Gurka, J. J., Li, J., and Bachmeier, A. S., 2005. Introducing the next generation advanced baseline imager on GOES-R. *Bulletin of the American Meteorological Society*, **1079**–**1096**.
- Stackhouse, P. W., Jr., Gupta, S. K., Cox, S. J., Mikovitz, J. C., and Chiaachio, M., 2002. New results from the NASA/GEWEX Surface Radiation Budget Project: evaluating El Nino effects at different scales. *11th Conference on Atmospheric Radiation*, *American Meteorological Society*, Ogden, UT, 3–7 Jun 2002.
- Stephens, G. L., et al., 2002. The CloudSat mission and the A-Train. *Bulletin of the American Meteorological Society*, **83**, 1771–1790, doi:http://dx.doi.org/10.1175/BAMS-83-12-1771.
- Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., Illingworth, A. J., O'Connor, E. J., Rossow, W. B., Durden, S. L., Miller, S. D., Austin, R. T., Benedetti, A., Mitrescu, C., and the ClouSat Science Team, 2002a. The CloudSat mission and the A-train: a new dimension of space-based observations of clouds and precipitation. *Bulletin of the American Meteorological Society*, **83**, 1771–1790.
- Stephens, G. L., Vane, D. G., Boain, R. J., and The ClouSat Science Team, 2002b. The Cloudsat mission and the A-Train: a new dimension of space-based observations of clouds and precipitation. *Bulletin of the American Meteorological Society*, **12**, 1171–1790.
- Stephens, G. L., Li, J., Wild, M., et al., 2012. An update on Earth's energy balance in light of the latest global observations. *Nature Geoscience*, **5**(10), 691–696, doi:10.1038/NCEO1580.
- Stuhlmann, R., Rieland, M., and Raschke, E., 1990. An improvement of the IGMK model to derive total and diffuse solar radiation at the surface from satellite data. *Journal of Applied Meteorology*, **29**, 586–603.
- Trenberth, K. E., Fasullo, J. T., and Kiehl, J., 2009. Earth's global energy. *Bulletin of the American Meteorological Society*, **90**, 311–323, doi:10.1175/2008BAMS2634.1.
- Valero, F. P. J., Pope, S. K., Bush, B. C., Nguyen, Q., Marsden, D., Cess, R. D., Simpson-Leitner, A. S., Bucholtz, A., and Udelhofen, P., 2003. Absorption of solar radiation by the clear and cloudy atmosphere during the Atmospheric Radiation Measurement Enhanced Shortwave Experiments (ARESE) I and II: observations and models. *Journal of Geophysical Research-Atmospheres*, **108**, doi:10.1029/2001DJ001384.
- Vardavas, I. M., and Taylor, F. W., 2007. Radiation and climate. *International Series of Monograph on Physics*, **138**, 512 pp.
- Wang, H., and Pinker, R. T., 2009. Shortwave radiative fluxes from MODIS: model development and implementation. *Journal of Geophysical Research*, **114**, D20201, doi:10.1029/2008JD010442.
- Whitlock, C. H., et al., 1995. First global WCRP shortwave surface radiation budget dataset. *Bulletin of the American Meteorological Society*, **76**, 905–922.
- Whitlock, C. H., Charlock, T. P., Staylor, W. F., Pinker, R. T., Laszlo, I., Ohmura, A., Gilgen, H., Konzelman, T., DiPasquale, R. C., Moats, C. D., LeCroy, S. R., Wielicki, N. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Louis Smith, G., and Cooper, J. E., 1996. Clouds and the Earth's Radiant Energy System (CERES): an earth observing system experiment. *Bulletin of the American Meteorological Society*, **77**, 853–868, doi:10.1175/1520-0477(1996)077.
- Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee, R. B., Louis Smith, G., and Cooper, J. E., 1996. Clouds and the Earth's Radiant Energy System (CERES): an Earth observing system experiment. *Bulletin of the American Meteorological Society*, **77**, 853–868, doi:10.1175/1520-0477(1996)077.
- Winker, D. M., Tackett, J. L., Getzewich, B. J., Liu, Z., Vaughan, M. A., and Rogers, R. R. (2012). The global 3-D distribution of tropospheric aerosols as characterized by CALIOP. *Atmospheric Chemistry and Physics. Discussion*, **12**, 24847–24893, doi:10.5194/acpd-12-24847-2012.
- Zhang, Y. C., Rossow, W. B., and Lacis, A. A., 1995. Calculation of surface and top of the atmosphere radiative fluxes from physical

- quantities based on ISCCP data sets. 1. Method and sensitivity to input data uncertainties. *Journal of Geophysical Research*, **100** (D1), 1149–1165, doi:10.1029/94JD02747.
- Zhang, Y.-C., Rossow, W. B., Lacis, A. A., Oinas, V., and Mishchenko, M. I., 2004. Calculation of radiative flux profiles from the surface to top-of-atmosphere based on ISCCP and other global datasets: refinements of the radiative transfer model and the input data. *Journal of Geophysical Research*, **109**, D19105, doi:10.1029/2003JD004457.
- Zhang, Y., Rossow, W. B., and Stackhouse, P. W., Jr., 2006. Comparison of different global information sources used in surface radiative flux calculation: radiative properties of the near-surface atmosphere. *Journal of Geophysical Research*, **111**, D13106, doi:10.1029/2005JD006873.
- Zhang, Y., Rossow, W. B., and Stackhouse, P. W., Jr., 2007. Comparison of different global information sources used in surface radiative flux calculation: radiative properties of the surface. *Journal of Geophysical Research*, **112**, D01102, doi:10.1029/2005JD007008.
- Zhang, T., Stackhouse, P. W. Jr., Gupta, S. K., et al., 2013. The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: a systematic quality control, production and application approach. *Journal of Quantitative Spectroscopy & Radiative Transfer*, **122**, Special Issue: 127–140, doi:10.1016/j.jqsrt.2012.10.004.
- Zhou, Y., Kratz, D. P., Wilber, A. C., Gupta, S. K., and Cess, R. D., 2007. An improved algorithm for retrieving surface downwelling longwave radiation from satellite measurements. *Journal of Geophysical Research*, **112**, D15102, doi:10.1029/2006JD008159.

SURFACE TRUTH

Christopher Ruf
Department of Atmospheric, Oceanic and Space Sciences,
University of Michigan, Ann Arbor, MI, USA

Synonyms

Ground truth

Definition and overview

The purpose of a scientific remote sensing instrument is to estimate geophysical properties of a target from indirect measurements of its electromagnetic radiation, propagation, and scattering. Geophysical properties, such as ocean surface temperature or atmospheric abundance of a certain molecule, are deduced from their electromagnetic properties. The procedure typically requires approximations, assumptions, and the use of measurements that contain random and systematic errors. As a result, the accuracy and precision of the geophysical estimates are usually validated by some independent means. Surface truth refers to some form of independent knowledge about the geophysical property being estimated. It is used to characterize the quality of the remote sensing data product, often using such statistical measures as the root-mean-square or the average difference between the remote sensing estimate and the surface truth.

Point-to-point comparisons

In most cases, surface truth comparisons with remote sensing estimates are made on a point-by-point basis. For example, a surface truth measurement may be compared to the remote sensing estimate made from a satellite as it passes overhead. Individual remote sensing estimates usually represent an instantaneous spatial average of the geophysical parameter. For surface imagers, the average is over the horizontal resolution spot size of the sensor. For atmospheric sounders, the average is over a resolution volume that includes both horizontal and vertical resolution boundaries. Different types of surface truth measurements are averaged in different ways. Surface truth for some ocean parameters – including [sea surface temperature](#), wind speed and direction, and sea level height – is often provided by buoys and tide gauges (Wentz et al., 2000; Mitchum, 1998; Goodberlet et al., 1989). Buoy measurements are made at a single point in space but can be averaged over time. Surface truth for atmospheric profiles of temperature and humidity is often provided by radiosondes, that is, instrumented weather balloons released from the ground (Buehler et al., 2004; Ruf et al., 1994). Radiosondes make point measurements along the line of ascension of the balloon during its ascent time, which typically lasts 30–60 min through the bulk of the troposphere. For both buoys and radiosondes, their spatial and temporal sampling characteristics do not match those of the remote sensing estimate. It is therefore not necessarily the case that perfect measurements by each should agree. Allowance for the sampling differences is made when they are compared.

Another class of point-to-point comparisons makes use of numerical weather forecast models such as those maintained by the US National Weather Service's National Center for Environmental Prediction (NCEP), the European Centre for Medium-Range Weather Forecasts (ECMWF), and the US Navy's Fleet Numerical Meteorology and Oceanography Center (FNMOC). These models produce global maps of atmospheric and surface parameters that are updated at least several times per day. The spatial resolution of the models is generally coarser than that of the remote sensing measurement, so there is once again a mismatch between their spatial sampling characteristics. In addition, the times at which a model is updated generally do not correspond exactly with the time when the remote sensing measurement is made, so temporal sampling is also mismatched. In spite of these differences, numerical forecast models are often used as surface truth. Relative to in situ sources of surface truth, such as buoys and radiosondes, they are particularly useful for assembling a large intercomparison database in a short period of time that is representative of the global variability of a particular geophysical parameter. One important limitation is the fact that these models tend to underestimate large deviations of the geophysical parameters from their mean values, and they generally perform less accurately in extreme weather conditions such as heavy precipitation and very high winds.

Statistical properties as surface truth

Some statistical properties of geophysical properties can also be used as a form of surface truth. To be useful, the property must be independently predictable, and reliable estimates of it must be obtainable from the available population of remote sensing measurements. One example is the global average wind speed and prevailing wind direction, both of which can be accurately predicted from either a large, globally distributed database of buoys or from a collection of numerical forecast models of the wind field (Frielich and Challenor, 1994; Ebuchi, 1999). Another example is the lower bound on microwave thermal emission by the ocean surface, which is determined by the temperature dependence of the emissivity of water (Ruf, 2000). These statistics tend to be very stationary over time and so are good candidates for statistical surface truth. If the statistics derived from surface truth and remote sensing do not agree, it is usually an indication that either the remote sensing instrument calibration is biased or the relationship between electromagnetic and geophysical properties that is assumed by the remote sensing estimation algorithm is in error. One important difference between point-to-point and statistical surface truth comparisons is that, with the statistical approach, the measurements with which the two statistics are derived need not coincide in space or in time. This can significantly simplify assembly of a satisfactory surface truth database, and it can ameliorate many of the effects of sampling differences between the surface truth and remote sensing measurements. However, these benefits come at the cost of a less direct type of comparison that does not as specifically quantify the performance of the remote sensing estimates as does the point-to-point method.

Conclusions

Surface truth represents knowledge about the geophysical properties being estimated by remote sensing that can be used to assess the quality of those estimates. The knowledge may result from an independent measurement of the geophysical property by some other means, such as a direct, in situ temperature or humidity probe, or it may result from statistical processing of a population of independent measurements. Surface truth is often used both to calibrate and validate a remote sensing data product and to characterize its quality after calibration is complete.

Bibliography

- Buehler, S. A., Kuvatov, M., John, V. O., Leiterer, U., and Dier, H., 2004. Comparison of microwave satellite humidity data and radiosonde profiles: a case study. *Journal of Geophysical Research*, **109**(D), 13103.
- Ebuchi, N., 1999. Relationship between directional distribution of scatterometer-derived winds and errors in geophysical model functions. *Journal of Advanced Marine Science and Technology Society*, **5**(1,2), 19.
- Frielich, M. H., and Challenor, P. G., 1994. A new approach for determining fully empirical altimeter wind speed model functions. *Journal of Geophysical Research*, **99**(C12), 25051.

- Goodberlet, M. A., Swift, C. T., and Wilkerson, J. C., 1989. Remote sensing of ocean surface winds with the special sensor microwave/imager. *Journal of Geophysical Research*, **94**, 14547.
- Mitchum, G., 1998. Monitoring the stability of satellite altimeters with tide gauges. *Journal of Atmospheric and Oceanic Technology*, **15**, 721.
- Ruf, C. S., Keihm, S. J., Subramanya, B., and Janssen, M. A., 1994. TOPEX/POSEIDON microwave radiometer performance and in-flight calibration. *Journal of Geophysical Research*, **99**(C12), 2491.
- Ruf, C. S., 2000. Detection of calibration drifts in spaceborne microwave radiometers using a vicarious cold reference. *IEEE Transactions on Geoscience and Remote Sensing*, **38**(1), 44.
- Wentz, F. J., Gentemann, C. L., Smith, D. K., and Chelton, D., 2000. Satellite measurements of sea surface temperature through clouds. *Science*, **288**, 847.

Cross-references

[Calibration and Validation](#)
[Geophysical Retrieval, Overview](#)

SURFACE WATER

Michael Durand
 School of Earth Sciences, The Ohio State University,
 Columbus, OH, USA

Definition

Remote sensing of surface water can be defined as collecting information about lakes, wetlands, floodplains, and rivers when viewed from above using electromagnetic radiation (Rees, 2001) or gravity field fluctuations (e.g., Crowley et al., 2006).

Introduction

Spaceborne and airborne instruments are used to characterize hydrologic features of surface water bodies: lakes, wetlands, floodplains, and rivers. Such remote sensing measurements are an increasingly important complement to traditional in situ surface water hydrology measurements. Alsdorf et al. (2007a) summarize science questions related to surface water and provide a detailed review of surface water remote sensing methods.

Primary quantities of interest for surface water quantity are the spatial extent of surface water bodies (inundated area, hereafter) and the vertical elevation of surface water bodies (water surface elevation, hereafter). From these primary quantities, changes in storage and hydrologic fluxes (e.g., river discharge) can be inferred. Primary quantities of interest for surface water quality relate to the biogeochemical constituents present in the water. This review of surface water remote sensing is organized around the following three broad fields: inundated area mapping, water surface elevation estimation, and water biogeochemical constituent characterization. For each of these three quantities, the remote sensing instruments and methods are explained, and some representative

applications of the *geophysical retrieval* (see entry [Geophysical Retrieval, Overview](#)) of surface water characteristics are described.

Inundated area mapping

The radiometric reflectivity and emissivity and radar backscatter behavior of liquid water allows for surface water hydrologic features to be distinguished from other geophysical entities at a number of wavelengths in the electromagnetic spectrum, which allows for the characterization of inundated area. For instance, the reflectance of clear water at visible and near-infrared wavelengths is less than that of most naturally occurring materials (Rees, 2001). Similarly, the microwave backscatter coefficient from open water is relatively low, permitting discrimination from land (Smith, 1997). Moreover, the passive microwave emissivity from inundated land is generally low, and the emissivity polarization difference is generally large (Prigent et al., 2007).

Both passive and active remote sensing instruments have been used to map inundated area for lakes, wetlands, floodplains, and rivers. For instance, Smith et al. (2005) assessed the total area and number of Siberian lakes using passive optical/infrared systems (see entry [Optical/Infrared, Radiative Transfer](#)): Landsat-1 in 1973 and the Russian RES URS-1 satellite in 1998. The comparison allowed for the quantification of a widespread decline in lake number and area in the 25 years between the two assessments. Prigent et al. (2007) present an excellent review of visible, infrared, and microwave passive methods and synthetic aperture radar (SAR) (see entry [Calibration, Synthetic Aperture Radars](#)) methods for characterizing inundated area for wetland studies. Prigent et al. (2007) obtained monthly estimates of global time series of inundated area at a 0.25° spatial resolution by merging disparate but complementary data streams: Advanced Very High Resolution Radiometer (AVHRR) visible and near-infrared (NIR) reflectances, the Special Sensor Microwave/Imager (SSM/I) passive microwave radiometer (see entry [Microwave Radiometers](#)) measurements, and radar scatterometer (see entry [Radar, Scatterometers](#)) measurements from the European Remote Sensing satellite.

Hess et al. (2003) used the Japanese Earth Resource Satellite-1 (JERS-1) SAR data to map inundated area at 100 m resolution in the Amazon Basin for two dates representing high water and low water. These inundated estimates have been useful in assessing the contribution of wetlands to biogeochemical processes in an area where obtaining in situ data is far from trivial. The Advanced Land Observing Satellite (ALOS) launched in January 2006 represents a significant improvement over JERS-1 and will permit more detailed floodplain and wetland characterization (Rosenqvist et al., 2007).

Many rivers show significant width (and, thus, inundated area) increase with increasing discharge, although this is not always the case for channelized rivers.

For instance, Smith and Pavelsky (2008) demonstrated that for the braided Lena River in Siberia, Moderate Resolution Imaging Spectroradiometer (MODIS) estimates of effective river width show strong correlation to river discharge. In addition to making inference on fluvial processes, inundated area is crucial for characterizing river flood events and evaluating flood inundation models (e.g., Horritt and Bates, 2001).

Water surface elevation estimation

Water surface elevations can be characterized by a variety of active (as opposed to passive) remote sensing instruments: laser altimetry, profiling radar altimetry, interferometric SAR, and swath radar altimetry. Laser altimeter systems function by emitting a pulse of visible or NIR radiation, detecting the echo of the pulse, and measuring the time between emission and detection. Traditional (profiling) radar altimetry (see entry [Radar, Altimeters](#)) works on much the same principles and provides a (spatially limited) profile of elevation estimates along the flight line (Elachi, 1987). Repeat-pass interferometric SAR (see entry [Ocean Applications of Interferometric SAR](#)) utilizes multiple SAR images to estimate elevation changes (e.g., Alsdorf et al., 2000). Imaging (swath) radar altimeters provide spatially continuous elevation estimates; SAR processing is very attractive for imaging radar altimeters (Elachi, 1987).

These methods are used for assessing surface water elevations in lakes, wetlands, floodplains, and rivers. For lakes and wetlands, water surface elevations represent a primary quantity of interest. For instance, Birkett (1994) used the Geosat profiling radar altimeter to characterize water surface elevations for lakes. Elevations of surface water stored in man-made reservoirs have been characterized using Shuttle Radar Topography Mission (SRTM) imaging SAR altimeter (Kiel et al., 2006). Yi et al. (2006) characterized surface water elevations in wetlands in Louisiana, USA, using TOPEX/POSEIDON profiling altimeter data. Changes in water surface elevation can be used to estimate changes in surface water storage.

River floodplain water elevations represent a combination of upstream hydrologic conditions and complex fluvial hydraulic processes. Alsdorf et al. (2007b) used JERS-1 SAR measurements to perform repeat-pass interferometric processing, which allowed for characterization of change in water elevation between the two satellite overpasses. This temporal change measurement provided new insight into the hydraulic complexity of the Amazon River floodplain. Limitations of the repeat-pass interferometric SAR include the following: (1) It does not characterize water height at a given time, but change in water height between two satellite overpasses, and (2) it requires flooded vegetation to return the radar pulse to the antenna.

Birkett et al. (2002) used TOPEX/POSEIDON profiling radar altimetry data to estimate water surface elevations for channelized rivers (as well as floodplains

and lakes) across the Amazon Basin. A major limitation of profiling radar altimetry for fluvial hydrology applications is the very limited spatial coverage and coarse spatial resolution. LeFavour and Alsdorf (2005) used SRTM water surface elevations to calculate river slope and then to estimate discharge along the Amazon River using simple flow hydraulics analysis. A limitation of the SRTM dataset is the temporal coverage: The entire mission was a total of 11 days in duration.

The Gravity Recovery and Climate Experiment (GRACE) maps changes in Earth's gravity field. These gravity changes are a result of several factors, including changes in surface water storage. GRACE thus differs from all other remote sensing methods mentioned herein, since the latter rely on observed electromagnetic radiation. For instance, Crowley et al. (2006) characterized water storage anomalies over the Congo River Basin for a 5 year period. A major limitation of GRACE is the coarse spatial resolution, which is on the order of hundreds of kilometers.

To overcome the limitations of existing remote sensing instruments, Alsdorf et al. (2007a) describe the need for an imaging (or wide swath) interferometric SAR altimeter to characterize water surface elevations and inundated area. The proposed satellite mission has been named Surface Water and Ocean Topography (SWOT) and has been recommended by the National Research Council Decadal Survey (NRC, 2007) to measure ocean topography as well as WSE over land; the proposed launch date time frame is between 2013 and 2016.

Water biogeochemical constituent characterization

In some cases, water biogeochemistry can be inferred from remote sensing when biogeochemical constituents alter the interaction of water with electromagnetic radiation. For example, turbid water generally has a higher spectral reflectance than clear water at visible and NIR wavelengths. Wang and Shi (2008) mapped a massive algae bloom in China's Lake Taihu using MODIS data. The MODIS spectral signature was used to quantify chlorophyll-a concentrations. Similarly, Chen et al. (2007) used the visible reflectances from the sea-viewing wide field-of-view sensor (SeaWiFS) to characterize a composite pollution index for the Pearl River estuary in China.

Summary

Spaceborne and airborne remote sensing measurements have been used to characterize inundated area, water elevations, and water biogeochemical constituents. Surface water is easily distinguished from surrounding terrain by active and passive sensors at a variety of wavelengths, permitting for the characterization of inundated area. Ranging instruments have been used to characterize water surface elevations, permitting the estimation of storage changes and river discharge, which are two crucial terms in the water balance. Remote sensing instruments have also been used to map water quality in terms of biogeochemical

constituents. The recent launch of ALOS PALSAR in January 2006 is expected to greatly increase the ability to effectively map floodplains and wetlands from space. The proposed launch of SWOT in 2013–2016 has the potential to allow for more detailed characterization of surface water storage changes and discharge than has so far been possible from a spaceborne platform.

Bibliography

- Alsdorf, D. E., Melack, J. M., Dunne, T., Mertes, L. A. K., Hess, L. L., and Smith, L. C., 2000. Interferometric radar measurements of water level changes on the Amazon flood plain. *Nature*, **404**, 174.
- Alsdorf, D. E., Rodríguez, E., and Lettenmaier, D. P., 2007a. Measuring surface water from space. *Reviews of Geophysics*, **45**, RG2002. 2006RG000197.
- Alsdorf, D. E., Bates, P., Melack, J., Wilson, M., and Dunne, T., 2007b. Spatial and temporal complexity of the Amazon flood measured from space. *Geophysical Research Letters*, doi:10.1029/2007GL029447.
- Birkett, C. M., 1994. Radar altimetry: a new concept in monitoring lake level changes. *Eos, Transactions American Geophysical Union*, **75**(24), 273.
- Birkett, C. M., Mertes, L. A. K., Dunne, T., Costa, M. H., and Jasinski, M. J., 2002. Surface water dynamics in the Amazon Basin: application of satellite radar altimetry. *Journal of Geophysical Research*, doi:10.1029/2001JD000609.
- Chen, C., Tang, S., Pan, Z., Zhan, H., Larson, M., and Jönsson, L., 2007. Remotely sensed assessment of water quality levels in the Pearl River estuary, China. *Marine Pollution Bulletin*, **54**, 1267.
- Crowley, J. W., Mitrovica, J. X., Bailey, R. C., Tamisiea, M. E., and Davis, J. L., 2006. Land water storage within the Congo basin inferred from GRACE satellite gravity data. *Geophysical Research Letters*, doi:10.1029/2006GL027070.
- Elachi, C., 1987. *Spaceborne Radar Remote Sensing: Applications and Techniques*. New York: IEEE Press.
- Hess, L. L., Melack, J. M., Novo, E. M. L. M., Barbosa, C. C. F., and Gastil, M., 2003. Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. *Remote Sensing of Environment*, **87**, 404.
- Horritt, M. S., and Bates, P. D., 2001. Predicting floodplain inundation: raster-based modeling versus the finite element approach. *Hydrological Processes*, **15**, 825.
- Kiel, B., Alsdorf, D. E., and LeFavour, G., 2006. Capability of SRTM C- and X-band DEM data to measure water surface elevations in Ohio and the Amazon. *Photogrammetric Engineering and Remote Sensing*, **72**(3), 1.
- LeFavour, G., and Alsdorf, D., 2005. Water slope and discharge in the Amazon River estimated using the shuttle radar topography mission digital elevation model. *Geophysical Research Letters*, doi:10.1029/2005GL023836.
- National Research Council, 2007. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. Washington, DC: The National Academies Press. 418 pp.
- Prigent, C., Papa, F., Aires, F., Rossow, W. B., and Matthews, E., 2007. Global inundation dynamics inferred from multiple satellite observations. *Journal of Geophysical Research*, doi:10.1029/2006JD007847.
- Rees, W. G., 2001. *Physical Principles of Remote Sensing*. Cambridge: Cambridge University Press.
- Rosenqvist, A., Shimada, M., Ito, N., and Watanabe, M., 2007. ALOS PALSAR: a pathfinder mission for global-scale monitoring of the environment. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 3307.

- Smith, L. C., 1997. Satellite remote sensing of river inundation area, stage, and discharge: a review. *Hydrological Processes*, **11**, 1427.
- Smith, L. C., and Pavelsky, T. M., 2008. Estimation of river discharge, propagation speed, and hydraulic geometry from space: Lena River, Siberia. *Water Resources Research*, doi:10.1029/2007WR006133.
- Smith, L. C., Sheng, Y., MacDonald, G. M., and Hinzman, L. D., 2005. Disappearing arctic lakes. *Science*, **308**, 1429.
- Wang, M., and Shi, W., 2008. Satellite-observed algae blooms in China's Lake Taihu. *Eos, Transactions American Geophysical Union*, **89**(22), 201.
- Yi, Y., Lee, H., Ibaraki, M., and Shum, C., 2006. Louisiana wetland monitoring using TOPEX/POSEIDON altimetry. *Eos, Transactions American Geophysical Union*, **87**(52), 1464.

Cross-references

[Geophysical Retrieval, Overview](#)
[Microwave Radiometers](#)
[Radar, Altimeters](#)
[Radar, Scatterometers](#)
[Radar, Synthetic Aperture](#)

TERRESTRIAL SNOW

Son V. Nghiem¹, Dorothy K. Hall², James L. Foster³ and Gregory Neumann¹

¹Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

²Cryospheric Sciences Laboratory, Code 615, NASA/Goddard Space Flight Center, Greenbelt, MD, USA

³Hydrological Sciences Laboratory, Code 617, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, MD, USA

Definitions

Among key parameters characterizing seasonal snow cover in cold land regions are snow area, extent, depth, water equivalent, accumulation onset, melt onset, ice layer, melt duration, and season length. Within an area $p(i,j)$ that has a fractional snow cover of $f_S(i,j)$, the actual area that is fully covered by snow is $s_A(i,j) = f_S(i,j) \cdot p(i,j)$ at a location determined by indices i and j on the Earth's surface. Then, the total snow area S_A is the summation of all $s_A(i,j)$. Depending on the sensitivity, accuracy, and resolution of a satellite sensor, the existence of snow cover in $p(i,j)$ can be detected when $s_A(i,j)$ is sufficiently large. The composition of all areas $p(i,j)$ where snow is detectable constitutes the snow extent S_E , which is smaller than S_A unless snow fully covers all snow areas $s_A(i,j)$.

Over area $p(i,j)$, the snow volume is calculated as $v_S(i,j) = d_a(i,j) \cdot p(i,j)$ where $d_a(i,j)$ is the average snow depth in area $p(i,j)$, including the snow-free fraction $[1 - f_S(i,j)]$ in $p(i,j)$ with zero snow depth. Rather than snow volume, it is more useful for hydrologists to use snow water equivalent (SWE), which is equal to the depth of water that ensues when the snow volume melts on $p(i,j)$. Let $\rho_a(i,j)$ be the average bulk density of snow and ρ_w the

water density, then $SWE(i,j) = d_a(i,j) \cdot \rho_a(i,j)/\rho_w$ is the average SWE over area $p(i,j)$.

The first snowfall in area $p(i,j)$ may be short lived because air and land temperatures are not cold enough to sustain it. As the landscape freezes, subsequent snowfalls may accumulate and contribute the snowpack during a snow season. The day when snow starts to consistently accumulate to build up the seasonal snowpack is called the snow accumulation onset date (T_S). During a snow season, an anomalous occurrence of snowmelt may create excessive meltwater, which refreezes in the snowpack, forming an ice layer that contains large icy objects (ice clumps, lens, columns, etc.).

As winter transitions into spring and then summer, snowmelt occurs more consistently and continuously (as opposed to isolated anomalous melt in winter such as that caused by warm air advection). Snowmelt onset date (T_M) is defined as the day when a consistent snowmelt starts to occur. Snow disappearance date (T_F) is the day when area $p(i,j)$ becomes snow-free as snow completely melts away. Then, snowmelt duration is $\Delta_M = T_M - T_F$ and snow season length is $\Delta_S = T_S - T_F$. These key snow parameters can be obtained using remote sensing data from multispectral visible infrared and microwave satellite sensors for each pixel $p(i,j)$.

Significance of snow cover

Climate change at high latitudes is strongly influenced by the albedo-temperature feedback process. The snow cover influences the global heat budget (Robinson and Kukla, 1985; Foster and Chang, 1993) because it affects the planetary albedo and the Earth's outgoing longwave radiation (Groisman et al., 1994). Snow is regarded as one of the key variables in global change monitoring (Walsh, 1991; Hall et al., 2005).

While large changes have been observed in many cold land environments (sea ice, ice sheet, glacier, tundra, and

permafrost), Bartlett et al. (2005) indicate no significant trends in the mean snow onset or duration over North America from 1950 to 2002, though their averaging approach over a hemispheric scale may suppress regional snow change signatures. Other studies (Chapin et al., 2005; Schwartz et al., 2006; Bonsal and Prowse, 2003) suggest that the Arctic snow season may have already decreased by more than 2.5 days per decade and that snow is melting earlier (Foster et al., 1992; Robinson et al., 1993; Frei and Robinson, 1999; Brown, 2000). This discrepancy illustrates a need for advanced satellite measurements of snow cover from local to regional and hemispheric scales.

Consistent with many climate change projections, significant changes in snow regimes are occurring throughout the Arctic. These changes have important consequences for humans, wildlife, and ecosystems. Recent changes in temperature and precipitation have altered the timing of the snow season and the snow-cover characteristics. The amount of snow (depth and *SWE*) has probably changed as well (Groisman and Davies, 2001), though this is a difficult quantity to measure accurately over large areas (Benson, 1982; Goodison et al., 1998; Yang et al., 2005; Liston and Sturm, 2004) compared to satellite measurements of snow extent. Furthermore, winter rain is likely more prevalent (Putkonen and Roe, 2003). It is critically important to comprehensively document these spatial and temporal changes with satellite data.

Moreover, satellite remote sensing of snow has immediate applications to snowmelt, river discharge, and flood forecasting. The timing and magnitude of river discharge in the Arctic is largely controlled by cold season snow mass storage and subsequent melt. In northern watersheds, snowmelt accounts for most of the annual river discharge and often produces the highest discharge rates of the year (Lo and Serreze, 2002; Bowling et al., 2000). Heavy snowstorms with rapid snowmelt in the spring of 1997 accounted for the “flood of the century” in the US northern Great Plains (Nghiem and Tsai, 2001), causing more than four billion dollars in flood-related damages (National Climatic Data Center, 1997). The ability to identify and track changes in snow regimes with increasing accuracy is an important asset for decision makers.

Seasonal snow extent

Satellite multispectral sensors have collected global data since the late 1960s, and algorithms for snow applications have been developed and improved. Satellite visible-band radiometers detect snow by its high albedo and characteristic patterns (Wiesnet and Matson, 1979) and by its spectral signature. An archive of quality-controlled National Oceanic and Atmospheric Administration (NOAA) weekly and now daily snow products is available from the Rutgers Global Snow Lab at Rutgers University (<http://climate.rutgers.edu/snowcover/>). The Earth Observing System (EOS) Terra and Aqua spacecraft, launched in 1999 and 2002, respectively, each carry an

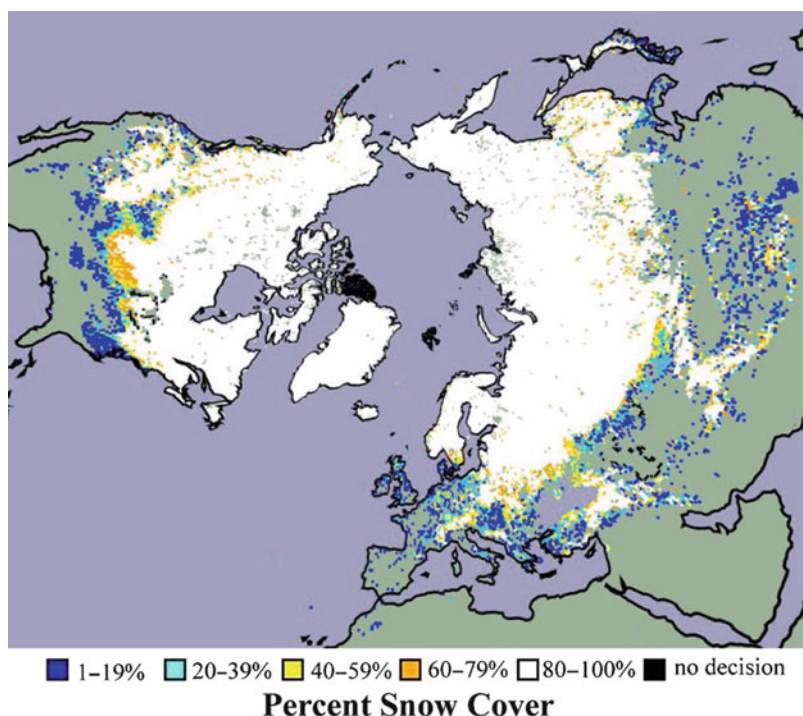
instrument called the Moderate Resolution Imaging Spectroradiometer (MODIS). MODIS has 36 channels covering visible to thermal infrared parts of the spectrum (0.4–14.4 μm) with a pixel size of 250 m–1 km and a swath width of 2,330 km, providing global coverage daily in most seasonally snow-covered regions.

MODIS data are used to obtain snow-cover products, including snow extent (S_E) and snow fraction (f_s), from automated algorithms at the Goddard Space Flight Center (GSFC) in Greenbelt, Maryland (<http://modis-snow-ice.gsfc.nasa.gov>). MODIS snow products are archived and distributed at the National Snow and Ice Data Center (NSIDC) in Boulder, Colorado. Global snow products are provided at 500 m, 5 km, or 25 km resolution as daily, 8 day composite, and monthly validated snow maps (Hall et al., 2002; Hall and Riggs, 2007). An example of a monthly, global MODIS snow-cover map is presented in Figure 1 for February 2004. The MODIS map shows snow extent from mid- to high latitudes over the Northern Hemisphere with smaller snow fraction at midlatitudes. Note that Western Europe (UK, France, Denmark, Germany, etc.) was either snow-free or did not have much snow cover compared to other regions at the same latitudes (Figure 1). At very high latitudes, such as in the Canadian Arctic Archipelago and Taymyr Peninsula in Siberia, there were difficulties in identifying snow due to winter darkness and cloud cover.

MODIS snow-cover maps complement existing hemispheric-scale snow maps that are available today (e.g., the NOAA maps (Ramsay, 1998)) and provide advances in spatial resolution and snow-cloud discrimination capabilities. The data archiving and distribution system provides free web-based access to the MODIS results. While MODIS has several technological advantages, the Advanced Very High Resolution Radiometers (AVHRR), with four to six channels (0.58–12.5 μm), have over three decades of data providing long-term measurements of seasonal snow extent (Foster et al., 2008a). Furthermore, active and passive microwave data can be used to supplement visible and near-infrared data over areas under clouds and darkness (Foster et al., 2008b).

Snow depth and snow water equivalent

The microwave brightness temperature emanating from the ground beneath a snow cover and from the snow itself is sensitive to snow depth, *SWE*, size of individual snow crystals, and other physical properties of snow. This interaction can be detected by spaceborne passive microwave radiometers. In the past three decades, considerable progress has been made in estimating regional and global snow depth and *SWE* using instruments such as the Scanning Multichannel Microwave Radiometer (SMMR) and the Special Sensor Microwave Imager (SSM/I). For more information on this topic, see results by Chang et al. (1987), Foster et al. (1997), Chen et al. (2001), and Pulliainen and Hallikainen (2001). Further progress has also been made in developing consistent multisensor



Terrestrial Snow, Figure 1 MODIS monthly snow-cover product (MOD10CM, Collection 5) for February 2004 showing snow extent and fraction of snow cover with a spatial resolution of 0.05° or about 5 km resolution. The color key is for fractional snow cover from 1 % to 100 %. Gray represents areas of persistent cloud cover according to the cloud mask.

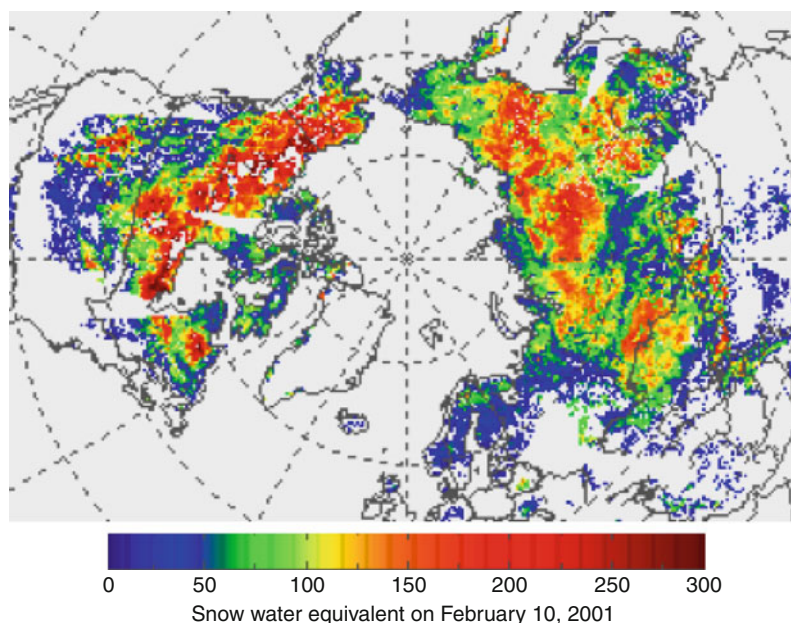
passive microwave *SWE* datasets for climate studies (Armstrong and Brodzik, 2001).

The Advanced Microwave Scanning Radiometer – EOS instrument (AMSR-E) on the Aqua satellite continued the heritage of earlier passive microwave Earth observations but included enhanced spatial resolution and an expanded range of frequency bands (6.925–89.0 GHz). The AMSR-E swath width was approximately 1,450 km. Research conducted in support of the implementation of AMSR-E for snowpack monitoring (Chang and Rango, 2000; Chang and Kelly, 2002; Kelly et al., 2003) has led to an automated algorithm for global *SWE*, at a pixel resolution of 25 km. The map in Figure 2 for February 10, 2001, shows an area of high *SWE* values in Labrador, Canada, which is consistent with the snow climatology in this region. However, due to the high sensitivity of the microwave response to varying snow properties, retrieval algorithms often produce large random estimation errors. Shallow or discontinuous snowpack, complex stratigraphy, large average snow grain size, or the presence of liquid water can adversely affect the algorithm performance and thus complicate the identification of changes in snow conditions.

Backscatter measurements from active microwave sensors complement the passive microwave data. Backscatter data from satellite scatterometers can also be useful for estimating snow depth and *SWE*. The SeaWinds scatterometer operating at Ku-band (13.4 GHz) aboard

the QuikSCAT satellite (QSCAT) was launched in June 1999. This sensor collects backscatter data with a resolution from 12 to 25 km over a swath of 1,400 km for the horizontal polarization (H) and 1,800 km for the vertical polarization (V). QSCAT already obtained global data for over a decade. Ku-band backscatter is sensitive to snow accumulation (Nghiem et al., 2001; Cline et al., 2004). For example, QSCAT backscatter signatures and snow parameters at a location in the Yellowstone National Park (44.65°N , 111.1°W) in Montana, USA, are shown in Figure 3. A regression analysis between QSCAT H backscatter and *SWE* data from the National Operational Hydrologic Remote Sensing Center (NOHRSC) National Snow Analysis (NSA) yields a positive correlation coefficient of 0.970, with a change of 108 mm per dB for *SWE* up to 200 mm. For *SWE* > 200 mm, the correlation is weaker and the relationship is less accurate because of attenuation effects in backscatter data and uncertainties in NOHRSC *SWE* results. Based on the rating curve derived from NOHRSC data (Nghiem et al., 2004), QSCAT data were used to estimate *SWE* values (bottom panel in Figure 3), corresponding to a range of snow density ratio (ρ_a/ρ_w) from 22 % to 56 % at Yellowstone.

The Yellowstone case represents the “scattering regime” where backscatter increases with snow depth and *SWE*. In the “attenuation regime” (see below “Anomalous melt events and ice layers”), backscatter decreases



Terrestrial Snow, Figure 2 Map of snow water equivalent (*SWE*) in millimeter (mm) derived from satellite microwave radiometer data (NASA/GSFC).

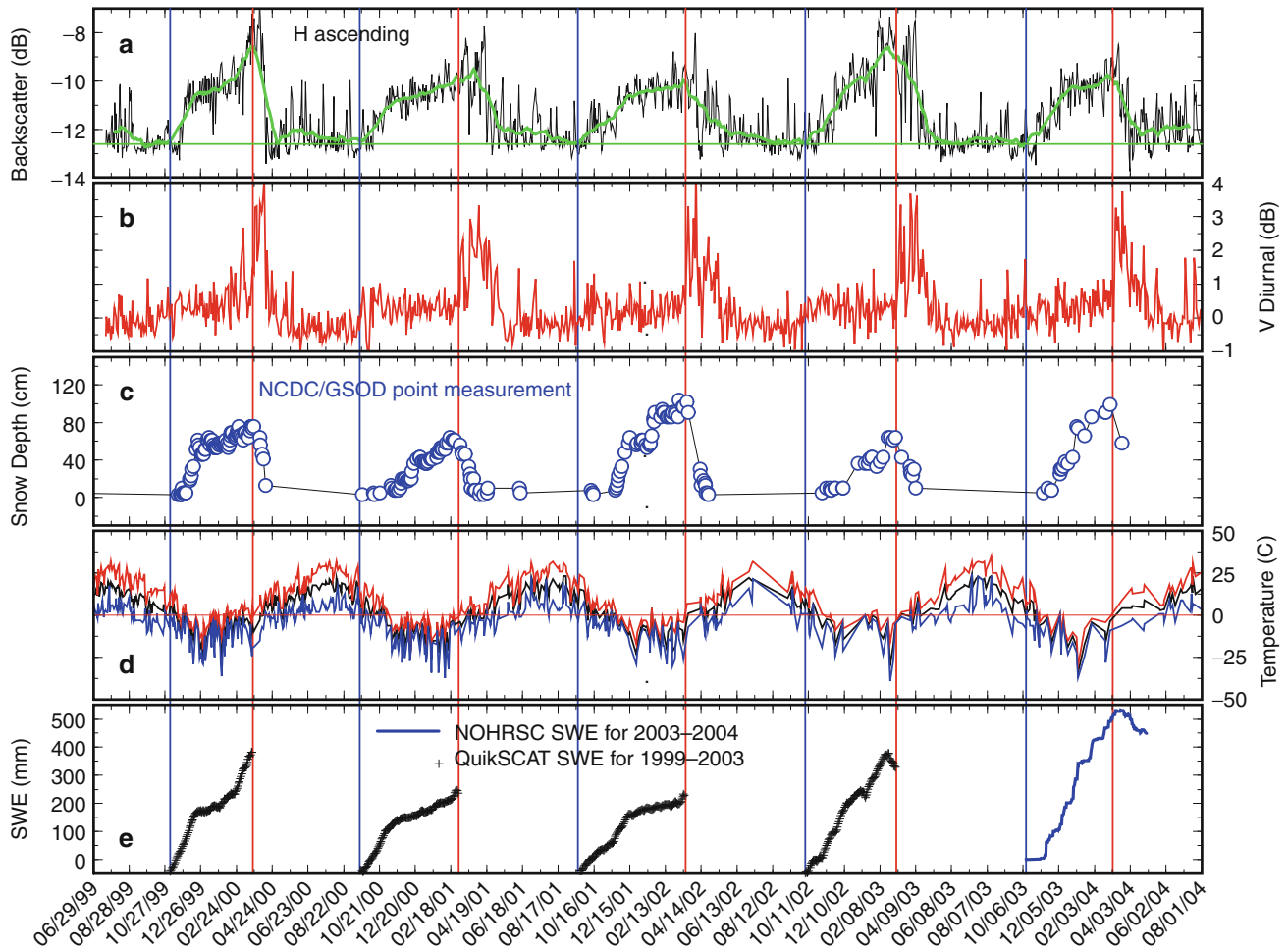
with snow accumulation. Both scattering and attenuation regimes can exist in a given area (Figure 4), so identifying the correct regime is important to the retrieval of snow depth and *SWE*. Temporal backscatter data can help improve results obtained from passive microwave snow observations. A new composite algorithm combining both active and passive data needs to be developed for more accurate *SWE* estimates.

Snowmelt pattern

With a wavelength of 2.2 cm (13.4 GHz), QSCAT backscatter is highly sensitive to snow wetness. Wave propagation and attenuation in an inhomogeneous medium like snow are characterized with a complex effective permittivity, which is determined by a mixture of air, ice grains, and liquid water (in the case of melting snow). The attenuation of Ku-band electromagnetic waves is controlled by the imaginary part of the snow effective permittivity. For dry snow, the imaginary part for pure dry water ice is in the order of $0.002\epsilon_0$ (Tiuri et al., 1984) where ϵ_0 is the permittivity of free space. In wet snow, liquid water (no salinity) has an imaginary part of about $38\epsilon_0$ (19,000 times larger than that of dry water ice) (Klein and Swift, 1977). Thus, a small amount of wetness can significantly change the imaginary part of the snow effective permittivity and, consequently, decrease the backscatter. This snow-scattering physics has been modeled by a number of researchers (Tsang et al., 1985; Ulaby et al., 1981; Nghiem et al., 1995). This sensitivity allows the backscatter to be used for snowmelt detection.

From experimental measurements with artificial piling of snow, a backscatter decrease of more than 5 dB is observed for 3 % wetness (e.g., Stiles and Ulaby, 1980). In a natural snow field, the Alaska Experiment (ALEX 1999) for snowmelt measurements was carried out in Ft. Wainwright, Alaska. In ALEX, a tower-mounted Ku-band scatterometer similar to QSCAT monitored a taiga snowpack, starting with freezing conditions and continuing through the snowmelt process until snow completely melted away (Nghiem et al., 2000). Taiga snow is a major snow class, covering about half of Alaska and Canada (Sturm et al., 1995). Results from ALEX show that diurnal backscatter can change as much as 15 dB (more than 30 times) between cold morning and warm afternoon when the heat makes more meltwater in the snowpack. These measurements confirm the high sensitivity of the backscatter to snowmelt.

The diurnal difference method was developed by Nghiem et al. (2001) to monitor the snowmelt process. The diurnal backscatter difference is a relative quantity between morning and afternoon measurements in half a day. QSCAT data from the early morning (t_a) in an ascending orbit pass and from late afternoon (t_p) in a descending pass are collocated for each day. The diurnal backscatter change is defined as the backscatter difference in the decibel (dB) domain as $\Delta\sigma_{VV} = \sigma_{VV}(t_p) - \sigma_{VV}(t_a)$ (Nghiem et al., 2001), where σ_{VV} is the V backscatter and all quantities are in dB. We use σ_{VV} to take advantage of the larger swath and coverage. In the case of snowmelt on the Greenland ice sheet, the criteria for the melt detection is based on $\Delta\sigma_{VV}$ greater than 1.8 dB for melt and less



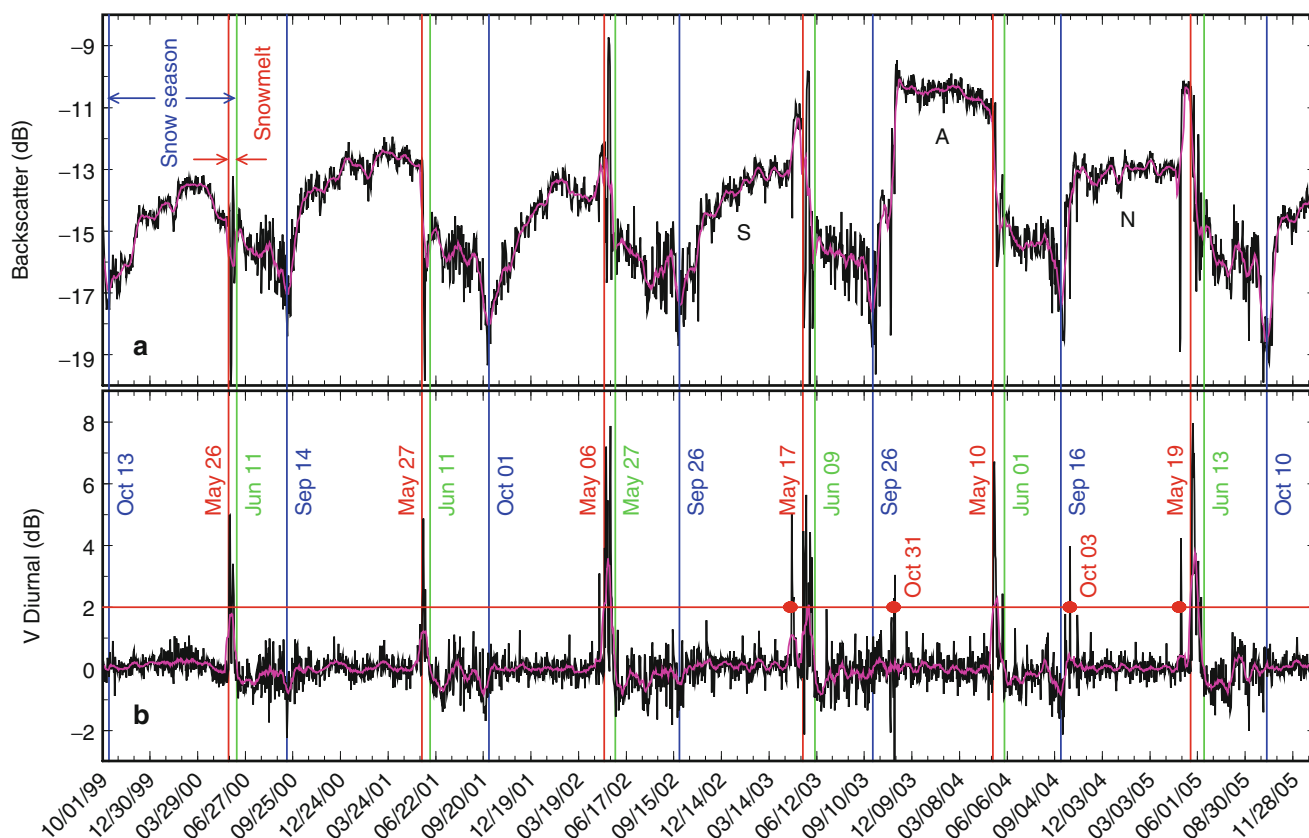
Terrestrial Snow, Figure 3 QSCAT signatures and snow parameters at Yellowstone (44.65 °N and 111.1 °W) in Montana, USA: (a) H backscatter from ascending orbits, (b) Diurnal V backscatter difference, (c) snow depth from the National Climatic Data Center (NCDC) Global Summary of the Day (GSOD) data, (d) minimum, average, and maximum air temperatures, and (e) SWE from QSCAT and from NOHRSC data. Snow depth and temperature data are from NCDC GSOD dataset. This area consists of 65 % coniferous forest, 23 % grasslands, 5 % woody savannas, and 4 % shrub lands (3 % unknown). Vertical blue lines indicate snow accumulation onsets, and red vertical lines mark snowmelt onsets.

than 1.0 dB for reduced melt or refreezing conditions (Steffen et al., 2004). This algorithm is adapted for snowmelt detection on land.

This method has several advantages: independence from the scatterometer long-term gain drift, independence from the cross-calibration between QSCAT and future satellite scatterometers, independence from absolute backscatter from different snow classes and snow conditions, detection of both snowmelt and refreezing, and daily coverage over most cold land regions. A unique feature of this method is that the diurnal backscatter difference stays near 0 dB before and after snowmelt. This feature holds for different snow classes under different physical conditions in different areas or regions. Therefore, a consistent and accurate method for snowmelt mapping can be applied over the continental scale. Because snowmelt is

a temporal process, each pixel is treated not as an independent measurement in time for a given snapshot, but as a part of an ongoing time-dependent process represented by a multiyear dataset for each pixel $p(i,j)$. The conglomerate of land pixels is analyzed to map snowmelt patterns at the continental scale.

The diurnal difference algorithm is applied to monitor snowmelt from mid-February to mid-June (Figure 5). The snowmelt map on March 21, 2007 (spring equinox), reveals a band of snow with reduced melt or refrozen conditions (light blue) stretching from the west of the Great Lakes toward the Pacific coast on North America. Below and adjacent to this reduced melt band was a stretch of land where the snowmelt process was completed (light brown). In Eastern Europe and Siberia, several snow regions can be observed at midlatitudes (red). Overall,



Terrestrial Snow, Figure 4 QSCAT signatures at Tunalik (70.2°N, 161.1°W) in Alaska, USA. The upper panel is for V backscatter and the lower panel is for diurnal V backscatter change. Vertical blue lines indicate snow accumulation onsets (T_S), red vertical lines mark melt onsets (T_M), and green lines are for snow departure dates (T_F). In the upper panel, horizontal blue arrows span the snow season length (Δ_S) and the red arrows mark the snowmelt period (Δ_M). The red dots on the horizontal red line (2 dB) in the QSCAT diurnal plot indicate anomalous melt events occurring in a snow season before melt onset. In the upper panel, symbol S denotes the scattering regime, A represents the attenuation regime, and N is for the neutral regime.

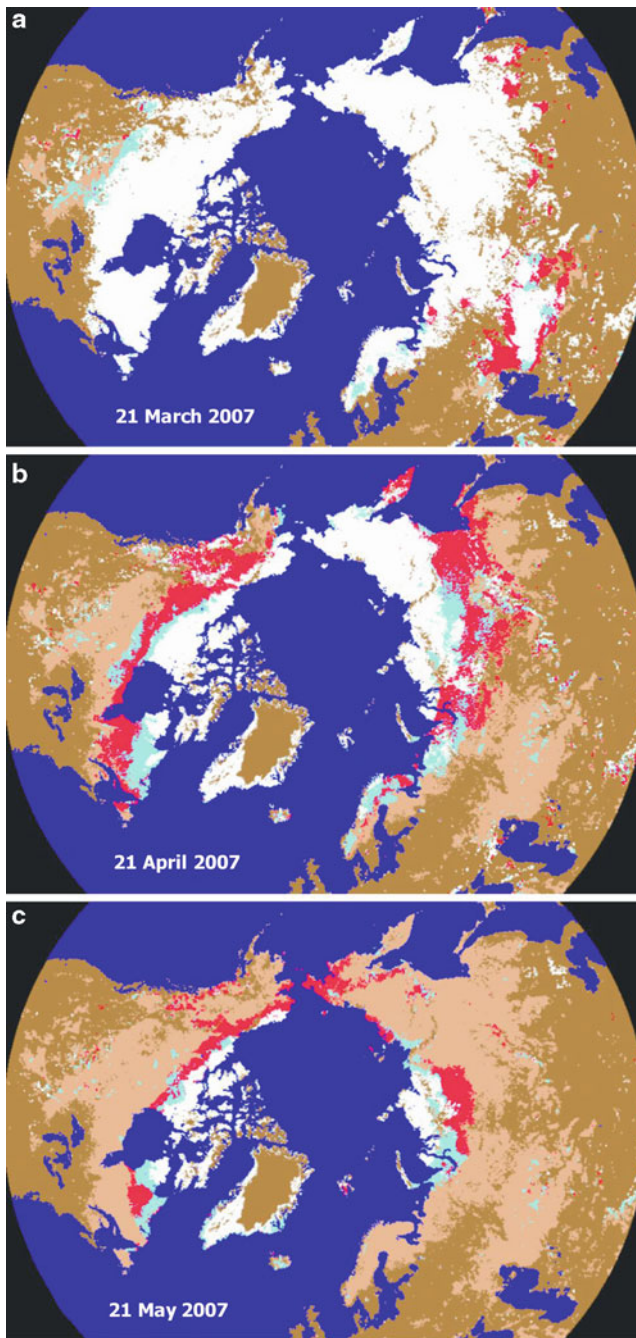
most of the Northern Hemispheric snowpack was in freezing conditions by the spring equinox. One month later (April 21, 2007), a very extensive melt band occurred at higher latitudes from Labrador across the Canadian Shield to Alaska, while snowmelt in Eastern Europe was completed and large areas of active melt were seen in eastern Siberia. By May 21, 2007 (a month before the summer solstice), most of the snowpack was melted except at very high latitudes in the Canadian Arctic, the Alaskan North Slope, and the Taymyr Peninsula in north Siberia. It is fortuitous that the QSCAT dataset covers this decade when significant changes have occurred, and this dataset is therefore invaluable for observations of changes in the Arctic snowmelt pattern.

Anomalous melt events and ice layers

Passive microwave emission is dependent on temperature and moisture in snow (Hallikainen et al., 2002). Melting

and refreezing of snow results in wet and dry snow metamorphism and changes in snow grain size (Colbeck, 1986; Colbeck et al., 1992). Refreezing of liquid water, from either melt or rain, can form an ice layer with large scatterers. These changes modify the passive microwave signature and may affect its sensitivity to snow accumulation.

For QSCAT backscatter, the scattering at Ku-band frequency follows the Rayleigh scattering law (Nghiem and Tsai, 2001), which dictates the scattering cross section to be proportional to the sixth power of grain size. Thus, QSCAT data can be used as a highly sensitive indicator for changes caused by snow metamorphism from processes such as melt and refreezing. For example, an anomalous event occurred at Tunalik on October 31, 2003 (Figure 4), when in situ meteorological data showed melt and rain, followed by refreezing. This event is characterized by a sharp peak in the QSCAT backscatter at the time indicated by the second red dot on the red horizontal line in the QSCAT diurnal plot in Figure 4. Collocated and



Terrestrial Snow, Figure 5 QSCAT snowmelt patterns on the 21st day of March, April, and May in 2007. White represents frozen snow, red is for active snowmelt, blue for snow with reduced melt or refrozen conditions, light brown for areas where the snowmelt process is completed, dark brown for land where QSCAT cannot detect snowmelt, and dark blue for water (oceans and lakes).

contemporaneous data from AMSR-E and SSM/I revealed that brightness temperature started to decrease on the same date; however, the minimum brightness temperature would not occur until 1 week later.

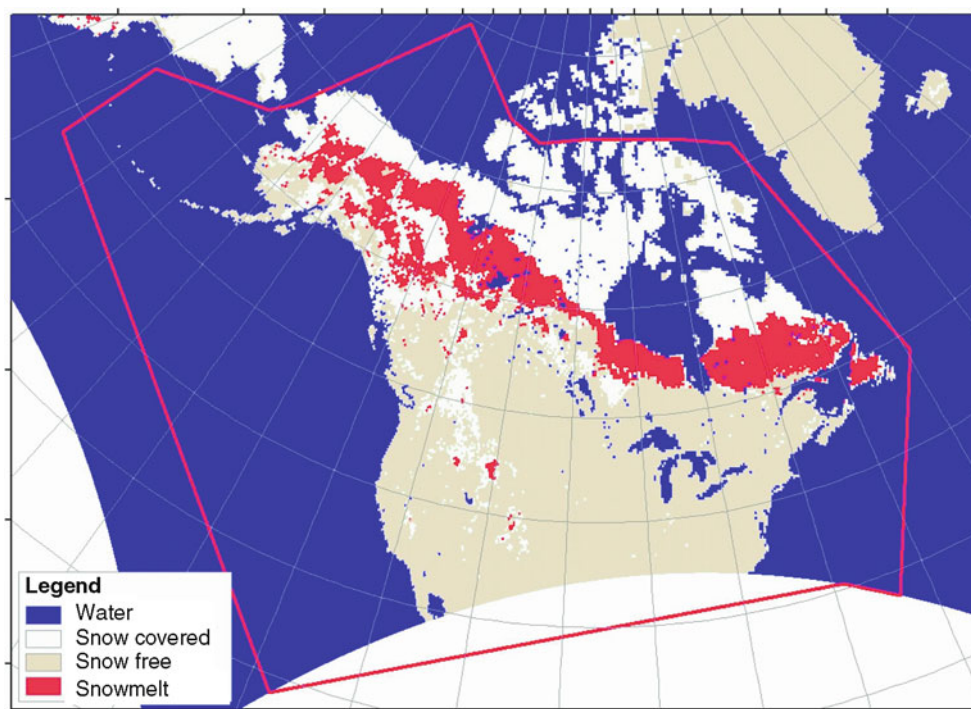
The sharp increase in QSCAT backscatter on October 31, 2003, and the decreasing trend in backscatter afterward indicate the formation of an ice layer in snow (Nghiem et al., 2005). Such detection allows a determination of the formation timing and extent of the ice layer. Moreover, QSCAT backscatter decreased as snow accumulated after the ice layer formation (October 31, 2003), indicating that the strong backscatter from the ice layer was attenuated by the subsequent overlaying snow accumulation (attenuation regime). In the previous years when there was no anomalous melt to form an ice layer, backscatter increased as snow accumulated from the start of each snow season, indicating that scattering increased with snow accumulation (scattering regime). In the snow season from September 2004 to May 2005, QSCAT backscatter did not show a clear trend with snow accumulation after the anomalous melt on October 3, 2004 (third red dot in the lower panel, Figure 4), indicating a neutral regime with a balance between scattering and attenuation effects. The detection of the switch between scattering and attenuation regimes or the identification of the neutral regime is useful for a better interpretation of passive microwave data for applications to *SWE* retrieval.

Timing of snow accumulation, melt onset, disappearance, and season length

Snow accumulation onset date (T_S), melt onset date (T_M), and disappearance date (T_F) are denoted, respectively, with blue, red, and green vertical lines in Figure 4 for the location at Tunalik in Alaska. Snow accumulation onset corresponds to the minimum in backscatter at the start of a snow season (Nghiem and Tsai, 2001). Snowmelt onset causes a large change in QSCAT diurnal signature consistently maintained in time without the long subsequent refreezing in the case of anomalous melt in winter. Snow disappearance occurs when backscatter is minimized and diurnal signature returns to a value near 0 dB. At Tunalik, snowmelt duration (Δ_M) varied from 15 to 25 days, while snow season length (Δ_S) changed as much as a month in the 6 years (1999–2005) of the QSCAT data. In 2002, snowmelt onset was the earliest, followed by the longest snow-free season (Figure 4).

Snowmelt typically results in an increase in grain size at the snow surface due to preferential melting of smaller grains or due to multigrain coalescence. MODIS has sufficient sensitivity to observe the spectral signatures resulting from the change in surface grain size and has been used to distinguish grain size categories (e.g., small, medium, and large) where there is continuous snow cover. A combination of AMSR-E, QSCAT, and MODIS signatures (satellite data collocated in space and closest in time) can better constrain the identification of snow onset, melt events, and snow season length.

Rain on snow-free land may have a similar diurnal signature in microwave data and thus cause a misidentification of snowmelt over such area. Here, MODIS data can help prevent the misidentification



Terrestrial Snow, Figure 6 Blended snow product (April 20, 2003) with a combination of snow extent (white) obtained from MODIS data and active snowmelt areas (red) derived from QSCAT data in North America (land inside the red boundary).

because of the capability of MODIS data to determine snow-covered and snow-free areas better than microwave data for thin snow areas.

Blended products

Because each type of satellite sensor, such as MODIS, AMSR-E, or QSCAT, has different sensitivities to different snow parameters, combinations of measurements from different sensors can be used to complement and improve the accuracy of snow measurement in a way that cannot be achieved by using a single sensor. Moreover, for many practical applications, combining different snow measurements from disparate products based on different satellite datasets with various resolutions, acquisition times, coverages, and formats is indisputably challenging for many users and decision makers. In this view, development of a snow mapping product that blends together different snow measurements is beneficial. Efforts to blend multiple snow datasets into a single, global, daily, and user-friendly product, such as the Air Force Weather Agency/National Aeronautics and Space Administration Snow Algorithm (ANSA), and to improve these products were supported in a research project (see Foster et al., 2008b for further details).

A simple example of a blended product is presented in Figure 6. This product combines snow extent measured by MODIS together with actively melting areas obtained from QSCAT data. The map of blended MODIS-QSCAT product (Figure 6) shows melt areas that consistently

occur within the limit of snow extent on April 20, 2003, in North America. An extensive melt band was observed across North America from Newfoundland, Canada, to Alaska, USA, between 50° and 70° latitudes. There are snow areas at lower latitudes or lower elevations that are not detected by QSCAT. This can be due to the low sensitivity of QSCAT backscatter to thin snow in such areas. Future developments for blended products will combine more snow products and will utilize more complex algorithms to account for temporal and spatial differences.

Summary

Satellite remote sensing of key parameters characterizing terrestrial snow cover, including snow extent, depth, water equivalent, snowmelt patterns, and seasonal snow dynamics (timing of accumulation onset, melt onset, disappearance, and length of snow season), have been presented. Data from optical infrared sensors (e.g., MODIS and AVHRR), passive microwave radiometers (e.g., SSM/I and, in the recent past, AMSR-E), and active microwave scatterometers (e.g., QSCAT) are used to measure snow parameters from local to regional and global scales on a daily to weekly basis. Techniques to blend different snow products into a single user-friendly global product are under development. Furthermore, different signatures corresponding to different scattering and attenuation regimes are identified based on snow physics. Such information will be valuable in the development of future satellite missions for global snow measurements, such as the

Cold Regions Hydrology High-resolution Observatory (CoReH₂O) Mission (Rott et al., 2007) and the Snow and Cold Land Processes (SCLP) Mission (National Research Council, 2007).

Acknowledgments

The research carried out at the Jet Propulsion Laboratory, California Institute of Technology, was supported by the National Aeronautics and Space Administration (NASA) Terrestrial Hydrology Program and by the US Air Force (USAF) under an agreement with NASA. The research at the NASA Goddard Space Flight Center was also supported by the USAF. The authors would like to thank Janet Y. L. Chien for preparing Figure 1 and D. Cline of NOAA NOHRSC for the NSA SWE data at Yellowstone.

Bibliography

- Armstrong, R. L., and Brodzik, M. J., 2001. Recent northern hemisphere snow extent: a comparison of data derived from visible and microwave satellite sensors. *Geophysical Research Letters*, **28**, 3673–3676.
- Bartlett, M. G., Chapman, D. S., and Harris, R. N., 2005. Snow effect on North American ground temperatures, 1950–2002. *Journal of Geophysical Research*, **110**, F03008, doi:10.1029/2005JF000293.
- Benson, C. S., 1982. *Reassessment of winter precipitation on Alaska's Arctic Slope and measurements on the flux of wind blown snow*. Geophysical Institute, University of Alaska Report UAG R-288, 26 pp. (Available from Geophysical Institute, University of Alaska, P.O. Box 757320, Fairbanks, AK 99775–7320).
- Bonsal, B. R., and Prowse, T. D., 2003. Trends and variability in spring and autumn 0°C-isotherm dates over Canada. *Climatic Change*, **57**, 341–358.
- Bowling, L. C., Lettenmaier, D. P., and Matheussen, B. V., 2000. Hydroclimatology of the Arctic drainage basin. In Lewis, L. (ed.), *The Freshwater Budget of the Arctic Ocean*. New York: Springer. Chap. 4.
- Brown, R. D., 2000. Northern hemisphere snow cover variability and change, 1915–1997. *Journal of Climate*, **13**(13), 2339–2355.
- Chang, A. T. C., and Kelly, R. E. J., 2002. *Description of snow depth retrieval algorithm for ADEOS II AMSR*. EORC Bulletin: Technical Report No. 9, ISSN: 1346–7913, pp. 70–78.
- Chang, A. T. C., and Rango, A., 2000. *Algorithm Theoretical Basis Document (ATBD) for the AMSR-E Snow Water Equivalent Algorithm, Version 3.1*. Greenbelt, MD: NASA/GSFC.
- Chang, A. T. C., Foster, J. L., and Hall, D. K., 1987. Microwave snow signatures (1.5 mm to 3 cm) over Alaska. *Cold Regions Science and Technology*, **13**(2), 153–160.
- Chapin, F. S., Sturm, M., Serreze, M. C., McFadden, J. P., Key, J. R., Lloyd, A. H., McGuire, A. D., Rupp, T. S., Lynch, A. H., Schimel, J. P., Beringer, J., Chapman, W. L., Epstein, H. E., Euskirchen, E. S., Hinzman, L. D., Jia, G., Ping, C. L., Tape, K. D., Thompson, C. D. C., Walker, D. A., and Welker, J. M., 2005. Role of land-surface changes in arctic summer warming. *Science*, **310**, 657–660.
- Chen, C. T., Nijssen, B., Guo, J., Tsang, L., Wood, A. W., Hwang, J., and Lettenmaier, D. P., 2001. Passive microwave remote sensing of snow constrained by hydrological simulations. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 1744–1756.
- Cline, D., Yueh, S. H., Nghiem, S. V., and McDonald, K., 2004. Ku-band radar response to terrestrial snow properties. *Eos Transactions, AGU*, 85(47), Fall meeting supplement, Abstract H23D–1149.
- Colbeck, S. C., 1986. Classification of seasonal snow cover crystals. *Water Resources Research*, **22**(9), 59S–70S.
- Colbeck, S., Akitaya, E., Armstrong, R., Gubler, H., Lafeyuille, J., Lied, K., McClung, D., and Morris, E., 1992. *The International Classification for Seasonal Snow on the Ground*. Hanover, NH: U.S. Army CRREL, p. 23.
- Council, N. R., 2007. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. Washington, DC: The National Academies Press.
- Foster, J. L., and Chang, A. T. C., 1993. Global snow cover. In Gurney, R. J., Parkinson, C. L., and Foster, J. L. (eds.), *Atlas of Satellite Observations Related to Global Change*. Cambridge: Cambridge University Press, pp. 361–370.
- Foster, J., Winchester, J., and Dutton, E., 1992. The date of snow disappearance on the Arctic tundra as determined from satellite, meteorological station and radiometric in situ observations. *IEEE Transactions on Geoscience and Remote Sensing*, **30**(4), 793–798.
- Foster, J. L., Chang, A. T. C., and Hall, D. K., 1997. Comparison of snow mass estimates from a prototype passive microwave snow algorithm, a revised algorithm and snow depth climatology. *Remote Sensing of Environment*, **62**, 132–142.
- Foster, J. L., Robinson, D. A., Hall, D. K., and Estilow, T. W., 2008a. Spring snow melt timing and changes over Arctic lands. *Polar Geography*, **31**(3–4), 145–157.
- Foster, J., Hall, D. K., Eylander, J., Riggs, G., Kim, E., Tedesco, M., Nghiem, S. V., Kelly, R. E. J., and Choudhury, B., 2008b. A new blended global snow product using visible, passive microwave, and scatterometer data, 12.1. In *Nineteenth Conference on Hydrology, 88th American Meteorological Society Annual Meeting*, New Orleans, LA.
- Frei, A., and Robinson, D. A., 1999. Northern hemisphere snow extent: regional variability 1972–1994. *International Journal of Climatology*, **19**(14), 1535–1560.
- Goodison, B. E., Louie, P. Y. T., and Yang, D., 1998. *WMO soil precipitation measurement intercomparison*. Final Report, WMO/TD 872. Geneva: World Meteorological Organization, 212 pp.
- Groisman, P. Y., and Davies, T. D., 2001. Snow cover and the climate system. In Jones, H. G., Pomeroy, J. W., Walker, D. A., and Hoham, R. W. (eds.), *Snow Ecology*. Cambridge: Cambridge University Press, pp. 1–44.
- Groisman, P. Y., Karl, T. R., Knight, R. W., and Stenichkov, G. L., 1994. Changes of snow cover, temperature, and radiative heat balance over the northern hemisphere. *Journal of Climate*, **7**(11), 1633–1656.
- Hall, D. K., and Riggs, G. A., 2007. Accuracy assessment of the MODIS snow-cover products. *Hydrological Processes*, **21**(12), 1534–1547, doi:10.1002/hyp. 6715.
- Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., and Bayr, K. J., 2002. MODIS snow-cover products. *Remote Sensing of Environment*, **83**, 181–194.
- Hall, D. K., Kelly, R. E. J., Foster, J. L., and Chang, A. T. C., 2005. Estimation of snow extent and snow properties. In Anderson, M. (ed.), *Encyclopedia of Hydrologic Sciences*. Chichester: Wiley, pp. 811–830. Chap. 55.
- Hallikainen, M., Halme, P., Takala, M., and Pulliainen, J., 2002. Effects of temperature and moisture of snow and soil on SSM/I response to snow. *IEEE Transactions on Geoscience and Remote Sensing*. In *Proceedings of International Geoscience and Remote Sensing Symposium*, **1**, 680–682.
- Kelly, R. E., Chang, A. T., Tsang, L., and Foster, J. L., 2003. A prototype AMSR-E global snow area and snow depth algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(2), 230–242.

- Klein, L. A., and Swift, C., 1977. An improved model for the dielectric constant of sea water at microwave frequencies. *IEEE Transactions on Antennas and Propagation*, **AP-25**(1), 104–111.
- Liston, G. E., and Sturm, M., 2004. The role of winter sublimation in the Arctic moisture budget. *Nordic Hydrology*, **35**(4), 325–334.
- Lo, R., and Serreze, M. C., 2002. Using snow water equivalent to understand the hydrology of the Arctic drainage system. *American Geophysical Union*, Fall meeting 2002, Abstract H51A-0762, San Francisco, CA.
- National Climatic Data Center, 1997. *The spring of 1997, reviewing four significant weather events*. Asheville, NC: National Environmental Satellite, Data, and Information Service. Technical Report, TR 97-03.
- Nghiem, S. V., and Tsai, W. Y., 2001. Global snow cover monitoring with spaceborne Ku-band scatterometer. *IEEE Transactions on Geoscience and Remote Sensing*, **39**(10), 2118–2134.
- Nghiem, S. V., Kwok, R., Yueh, S. H., Kong, J. A., Tassoudji, M. A., Hsu, C. C., and Shin, R. T., 1995. Polarimetric scattering from layered media with multiple species of scatterers. *Radio Science*, **30**(4), 835–852.
- Nghiem, S. V., Sturm, M., Perovich, D. K., Tsai, W., Neumann, G., Taras, B., and Elder, B., 2000. Large-scale snow cover monitoring with Sea Winds/QuikSCAT scatterometer. *Eos Transactions, AGU*, **81**(48), Fall meeting supplement, Abstract H62G-04.
- Nghiem, S. V., Steffen, K., Kwok, R., and Tsai, W.-Y., 2001. Detection of snow melt regions on the Greenland ice sheet using diurnal backscatter change. *Journal of Glaciology*, **47**(159), 539–547.
- Nghiem, S. V., Yueh, S. H., Cline, D., Sturm, M., and Neumann, G., 2004. Ku-band radar signature and snow accumulation inversion. In *NASA Terrestrial Hydrology Meeting*, Maryland.
- Nghiem, S. V., Steffen, K., Neumann, G., and Huff, R., 2005. Mapping of ice layer extent and snow accumulation in the percolation zone of the Greenland ice sheet. *Journal of Geophysical Research*, **110**, F02017, doi:10.1029/2004JF00234.
- Pulliainen, J., and Hallikainen, M., 2001. Retrieval of regional snow water equivalent from space-borne passive microwave observations. *Remote Sensing of Environment*, **75**, 76–85.
- Putkonen, J., and Roe, G., 2003. Rain-on-snow events impact soil temperatures and affect ungulate survival. *Geophysical Research Letters*, **30**(4), 1188, doi:10.1029/2002GLO16326.
- Ramsay, B., 1998. The interactive multisensor snow and ice mapping system. *Hydrological Processes*, **12**, 1537–1546.
- Robinson, D. A., and Kukla, G., 1985. Maximum surface albedo of seasonally snow covered lands in the northern hemisphere. *Journal of Climate and Applied Meteorology*, **24**, 402–411.
- Robinson, D. A., Dewey, K. F., and Heim, R. R., 1993. Global snow-cover monitoring: an update. *Bulletin of the American Meteorological Society*, **74**(9), 1689–1696.
- Rott, H., Cline, D., Nagle, T., Pulliainen, J., Rebhan, H., and Yueh, S., 2007. CoRe-H2O – a dual frequency SAR mission for hydrology and climate research. In *Proceedings of International Geoscience and Remote Sensing Symposium*. Barcelona, pp. 1204–1206.
- Schwartz, M. D., Rein, A., and Aasa, A., 2006. Onset of spring starting earlier across the northern hemisphere. *Global Change Biology*, **12**, 343–351, doi:10.1111/j.1365-2486.2005.01097.
- Steffen, K., Nghiem, S. V., Huff, R., and Neumann, G., 2004. The melt anomaly of 2002 on the Greenland ice sheet from active and passive microwave satellite observations. *Geophysical Research Letters*, **31**(20), L20402, doi:10.1029/2004GL020444.
- Stiles, W. H., and Ulaby, F. T., 1980. The active and passive microwave response to snow parameters. I. Wetness. *Journal of Geophysical Research*, **85**(C2), 1037–1044.
- Sturm, M., Holmgren, J., and Liston, G., 1995. A seasonal snow cover classification system for local to global applications. *Journal of Climate*, **8**(5), 1261–1283.
- Tiuri, M. E., Sihvola, A. H., Nyfors, E. G., and Hallikainen, M. T., 1984. The complex dielectric constant of snow at microwave frequencies. *IEEE Journal of Oceanic Engineering*, **OE-9**(5), 377–382.
- Tsang, L., Kong, J. A., and Shin, R. T., 1985. *Theory of Microwave Remote Sensing*. New York: Wiley.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1981. *Microwave Remote Sensing: Active and Passive*. Massachusetts: Artech House.
- Walsh, J. E., 1991. Operational satellites and the global monitoring of snow and ice. *Global and Planetary Change*, **90**(1–3), 219–224.
- Wiesnet, D. R., and Matson, M., 1979. The satellite-derived northern hemisphere snowcover record for the winter of 1977–78. *Monthly Weather Review*, **107**, 928–933.
- Yang, D., Kane, D., Zhang, Z., Legates, D., and Goodison, B., 2005. Bias-corrections of long-term (1973–2004) daily precipitation data over the northern regions. *Geophysical Research Letters*, **32**, 119501, doi:10.1029/2005GL02405.

Cross-references

[Cryosphere, Climate Change Effects](#)
[Cryosphere, Climate Change Feedbacks](#)
[Cryosphere, Measurements and Applications](#)
[Data Archival and Distribution](#)
[Microwave Radiometers](#)
[Radars](#)
[Snowfall](#)

THERMAL RADIATION SENSORS (EMITTED)

Simon Hook
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Synonyms

Thermal infrared sensors

Definitions

Radiant energy. The energy of electromagnetic waves or sometimes of other forms of radiation. The quantity of radiant energy may be calculated by integrating radiant flux (or power) with respect to time, and, like all forms of energy, its SI unit is the joule. The term is used particularly when radiation is emitted by a source into the surrounding environment (Source: Wikipedia).

Thermal radiation. Electromagnetic radiation emitted from the surface of an object which is due to the object's temperature (Source: Wikipedia).

Infrared radiation. Electromagnetic radiation whose wavelength is longer than that of visible light but shorter than that of terahertz radiation and microwaves (Source: Wikipedia).

Sensor: Device that measures a physical quantity and converts it into a signal which can be read by an observer or by an instrument (Source: Wikipedia).

Introduction

In this entry, we consider sensors that remotely measure the emitted thermal radiation from the surface of the Earth and the associated atmosphere at wavelengths between 3 and 14 μm . These measurements are typically made from three vantage points: (1) in situ measurements, where the sensor is placed within a few meters of the surface; (2) airborne measurements, where the sensor is mounted on an aircraft flying at 100 m to many kilometers above the surface of the Earth but still within the Earth's atmosphere; and (3) spaceborne measurements, where the instrument is on a satellite platform orbiting the Earth. The emitted radiation that is sensed depends on the temperature and composition of the surface and the atmosphere and thereby provides a means for measuring changes in composition or temperature of the surface and/or atmosphere. These same sensor types are used to measure emitted thermal radiation from other planetary bodies over extended wavelength regions and can provide additional information on the composition of the surface depending on the composition of the overlying atmosphere. On Earth, within the 3–14 μm wavelength region, there are two atmospheric “windows.” These are regions where the energy from the surface is not as strongly blocked by the atmosphere; these are referred to as the mid infrared (MIR) and thermal infrared (TIR). In this entry, we consider the MIR window to extend between 3 and 5 μm and the TIR window between 8 and 12 μm . The terms MIR and TIR are used loosely and therefore it is best to define the wavelength region that is being considered when the term is first used. Since the signal from the surface of the Earth is strongest in the MIR and TIR, sensors that measure the emitted radiation in these wavelength regions are typically used to study the surface of the Earth. In contrast, studies of the Earth's atmosphere will typically make measurements over the entire wavelength region since the effect of gases within the Earth's atmosphere is pervasive throughout the region albeit stronger at selected wavelengths.

Remote measurement of emitted radiation

Remote sensing within the 3–14 μm wavelength region is based on the measurement of the radiance emitted from the surface and modified by the atmosphere; its basic goal is to recover the kinetic temperature and emissivity from the measured at-sensor radiance of the surface and/or the atmosphere. The at-sensor radiance (L_s) for a given wavelength (λ) in this wavelength region, excluding any reflected solar contribution which occurs below 5 μm , can be written as

$$L_{s\lambda} = [\varepsilon_\lambda L_{bb\lambda}(T) + (1 - \varepsilon_\lambda) L_{sky\lambda}] \tau_\lambda + L_{atm\lambda} \quad (1)$$

where ε_λ is the surface emissivity at wavelength λ , $L_{bb\lambda}(T)$ is the spectral radiance from a blackbody at surface temperature T , $L_{sky\lambda}$ is the sky radiance (spectral downwelling radiance incident upon the surface from the atmosphere), τ_λ is the transmittance (spectral atmospheric transmission), and $L_{atm\lambda}$ is the path radiance (spectral upwelling radiance from atmospheric emission and scattering that reaches the sensor).

The effect of τ_λ is to reduce the amount of ground-emitted radiance measured at the sensor; $L_{atm\lambda}$ adds a component unrelated to the ground and $L_{sky\lambda}$ serves to reduce the spectral contrast since the deeper an emissivity feature, the more it is “filled in” by reflected downwelling radiation due to Kirchhoff's Law. Any algorithm that is used to produce a fundamental geophysical parameter such as surface radiance, temperature, or emissivity must compensate for these atmospheric effects.

Once the atmospheric effects have been removed, the surface radiance is given by

$$L_\lambda = \varepsilon_\lambda L_{BB\lambda}(T) \quad (2)$$

Or

$$L_{BB\lambda} = \frac{C_1}{\lambda^5 \pi \left[e^{\left(\frac{C_2}{\lambda T}\right)} - 1 \right]} \quad (3)$$

where

L_λ = emitted radiance

$L_{BB\lambda}$ = blackbody radiance ($\text{Wm}^{-2} (\text{m} - \Delta\lambda)^{-1} \text{sr}^{-1}$)

λ = wavelength of channel (m)

T = temperature of blackbody (Kelvin)

C_1 = first radiation constant = $3.7415 \times 10^{-16} \text{ (Wm}^2\text{)}$

C_2 = second radiation constant = $0.0143879 \text{ (m deg)}$

ε_λ = surface emissivity (0–1)

The surface temperature (T) is not an intrinsic property of the surface; it varies with external factors such as the irradiance history and meteorological conditions. Conversely, emissivity is a characteristic of the material making up the surface and is independent of the temperature. For most terrestrial surfaces, $200 \text{ K} < T < 340 \text{ K}$, although a more restricted range of $\sim 270\text{--}310 \text{ K}$ probably brackets most of the temperatures outside the polar regions and deserts. The emissivities for most natural surfaces in MIR and TIR wavelength typically range vary $\sim 0.7\text{--}1$ (Prabhakara and Dalu, 1976) with surfaces with emissivities < 0.85 are largely restricted to arid and semiarid regions.

Equation 3 describes the radiance from a homogenous, isothermal surface. However, many surfaces at the resolution of current instruments consist of several materials at different temperatures and emissivities.

Airborne and spaceborne sensors

From the standpoint of most researchers, the ideal instrument would be one that provided measurements of the emitted radiation at very small wavelength increments and was able to resolve differences in energy at these

Thermal Radiation Sensors (Emitted), Table 1 Characteristics of the main emitted radiation detector types used in airborne and spaceborne instruments today

Detector type	Quantum efficiency	Operating temp	Residual nonuniformity	Response time	Cutoff wavelength ^a
HgCdTe	>80 %	40–80 K	High	Fast	12 μm
QWIP	30 %	40 K	Very low	Fast	12 μm
Uncooled Microbolometers	N/A	Room temp	Low	Slow	Flat response

Land surface temperature (LST) and emissivity are key variables for explaining the biophysical processes which govern the balances of water and energy at the land surface. This table summarizes the science requirements for the LSTE-ESDR

wavelength increments over very small distances on the surface of the Earth. Unfortunately no technology currently exists that meets these requirements, and as a result, most sensors provide data at very small wavelength increments (high spectral resolution) from large areas of the Earth's surface (low spatial resolution) or data at low spectral resolution over small distances on the Earth's surface (high spatial resolution). These trade-offs are necessary so the detector(s) in the sensor can receive sufficient radiation to enable measurement. There are three main detector types in operation today. These are Mercury Cadmium Telluride (MCT) detectors, microbolometers, and Quantum Well Infrared Photodetectors (QWIPs). MCT detectors and microbolometers are more widely used than the newer QWIP technology. Some of the characteristics of these detectors are summarized in Table 1. More detail on the types of detectors is available in Miles and Chow (1996). Microbolometer detectors do not require cooling which can provide a large power saving (cooling takes power), but they respond slowly. As a result, they are used with instruments that can either stare and integrate the radiation over long periods of time or, if used in an application that requires rapid imaging, provide measurements with low spatial and spectral resolution. MCT detectors require cooling but that cooling can be either passive (radiative coolers) or active (mechanical coolers). Radiative coolers are more desirable since they require less power but typically can only reach temperatures around 80 K. At these temperatures, either high spectral resolution and low spatial resolution measurements or low spectral resolution and high spatial resolution measurements are possible but not both, that is, high spatial and high spectral resolution. In order to reach the temperatures required for both high spatial and high spectral resolution, the detector needs to be cooled to temperatures around 40 K, typically requiring active cooling. Within the last decade, spaceborne mechanical coolers have become more available and there are two common types: either sterling cycle or pulse tube and modern sensors that push the current measurement capability use mechanically cooled MCT or QWIP detectors. All three types of detectors are available as one- or two-dimensional arrays. Arrays allow multiple points on the surface to be imaged at once at several wavelengths which provide more signal to each detector (longer integration time) when the time available to make the measurement is short and

consequently detector arrays are widely used today. QWIP detector arrays are more uniform than MCT arrays, which are much more desirable to enable consistent quantitative detection. These different trade-offs have the result that future high spectral and spatial resolution airborne and spaceborne sensors, where the time available to make the measurement is short, will likely use mechanically cooled detector arrays, either MCT or QWIPs. The so-called megapixel arrays ($1,000 \times 1,000$ elements) are now available for both types of detector. Other detector types continue to be developed which may replace MCT and QWIPs, but the adoption of new detector types such as strained layer super lattice detectors tends to be very gradual.

Table 2 provides a summary of a selection of the currently available spaceborne sensors and plans for future spaceborne sensors. These sensors represent the current and planned spaceborne state of the art however; they do not include any high spatial and high spectral resolution sensors. Such sensors do exist but currently only operate from airborne platforms. Airborne platforms are ideal for developing new spaceborne sensors since they provide a large platform with ample power and ready access to the instrument in flight. Many airborne instruments provide example datasets for future spaceborne instruments as well as more detailed spectral and spatial information as spaceborne instruments, allowing the performance of the spaceborne instruments to be validated.

Calibration and validation

Most airborne and spaceborne emitted radiation sensors include some form of onboard calibration. Typically this involves one or more blackbodies. A blackbody has an emissivity of 1.0, and therefore, the radiation emitted from it depends entirely on its temperature. This known temperature is used as a reference to bracket the scene temperatures together with a space look which also provides a blackbody measurement. A typical example of a spaceborne instrument is the Moderate Resolution Imaging Spectrometer (MODIS) which acquires data at several wavelength intervals (channels) between 3 and 14 μm (Barnes et al., 1998; Salomonson et al., 1989). MODIS scans 55° from nadir and provides daytime and nighttime imaging of any point on the Earth every 1–2 days with a continuous duty cycle. MODIS data are quantized in

Thermal Radiation Sensors (Emitted), Table 2 Examples of current and planned spaceborne emitted thermal radiation sensors with their spatial and spectral resolution ranges

Subproduct	Spatial resolution	Temporal resolution	Current data sources	Future data source
Global	10–20 km	Hourly	AIRS GOES MSG	CrIS GOES MSG
Regional	1–5 km	2–4 times daily	MODIS AVHRR ATSR	VIIRS AVHRR ATSR
Local	30–100 m	Once every 8–16 days	ASTER Landsat	

Atmospheric Infrared Radiation Sounder (AIRS), Geostationary Operational Environmental Satellites (GOES), Meteosat Second Generation (MSG), Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Very High Resolution Radiometer (AVHRR), Along Track Scanning Radiometer (ATSR), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Cross-track Infrared Sounder (CrIS), Visible Infrared Imager/Radiometer Suite (VIIRS)

12 bits and have a spatial resolution of ~ 1 km at nadir. They are calibrated with a cold space view and full aperture blackbody viewed before and after each Earth view (Salomonson et al., 1989; Barnes et al., 1998). There are many different types of blackbody; MODIS, for example, uses a flat blackbody with a V-groove. Flat blackbodies with a V-groove are often used instead of cone blackbodies since they occupy less physical space. However, from a scientific standpoint, the most accurate blackbodies are cone blackbodies. These are used, for example, on the European Along Track Scanning Radiometer series of instruments. Great care is taken to calibrate emitted thermal radiation sensors used for research such as the two sensors mentioned. This allows measurements from the sensors made over long periods of time (decades) to be tracked to observe small changes in incoming radiation and enable a variety of research studies such as climate change. While onboard calibration is essential, an independent measurement is needed to ensure that the onboard system is working correctly. This independent measurement is used to validate the onboard calibration system. Typically the validation measurement involves measuring the radiation from a sensor close to the ground or in an aircraft and then propagating that measurement through the Earth's atmosphere with a radiative transfer model (Berk et al., 1989) to obtain the radiance that the satellite instrument should observe and comparing it with what the satellite instrument measures. In the last decade a validation site was established at Lake Tahoe CA/NV, USA, where these measurements are made on a continuous basis and data from the site have been used to validate several sensors including ASTER, MODIS, Landsat (TM and ETM), and ATSR (Barsi et al., 2003; Hook et al., 2003, 2005, 2006; Tonooka et al., 2005).

Conclusion

Over the last few decades, multiple emitted radiation sensors have been launched into orbit around the Earth as well as other planetary bodies. Data from these sensors have

been used in a wide variety of studies from mapping the surface composition of a region to studying climate change and forecasting the weather. The instruments that have been launched were typically based on prototypes deployed from aircraft. Our ability to make even higher spectral and spatial measurements from space is limited by the available technology. As the technology evolves, so does the measurement capability and science that the sensors can be used to address. One area of concern is the lack of future sensors capable of making high spatial resolution thermal infrared measurements such as those currently provided by ASTER and Landsat (Yamaguchi et al., 1998; Barsi et al., 2003). Currently the research community is calling for such measurements, but it remains to be seen whether such measurements will be available in the future.

Acknowledgments

The research described in this entry was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Bibliography

- Barnes, W. L., Pagano, T. S., and Salomonson, V. V., 1998. Pre-launch characteristics of the moderate resolution imaging spectroradiometer (MODIS) on EOS AM1. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 1088–1100.
- Barsi, J. A., Schott, J. R., Palluconi, F. D., Helder, D. L., Hook, S. J., Markhum, B. L., Chander, G., and O'Donnell, E. M., 2003. Landsat TM and ETM + thermal band calibration. *Canadian Journal of Remote Sensing*, **29**(2), 141–153.
- Berk, A., Bernstein, L. S., and Robertson, D. C., 1989. *MODTRAN: A Moderate Resolution Model for LOWTRAN 7*. Technical Report GL-TR-89-0122. Bedford, MA: Geophysics Laboratory.
- Hook, S. J., Prata, A. J., Alley, R. E., Abtahi, A., Richards, R. C., Schladow, S. G., and Pálmarsson, S. Ó., 2003. Retrieval of lake bulk-and skin-temperatures using along track scanning radiometer (ATSR) data: a case study using Lake Tahoe, CA. *Journal of Atmospheric and Oceanic Technology*, **20**(2), 534–548.

- Hook, S. J., Clodius, W. B., Balick, L., Alley, R. E., Abtahi, A., Richards, R. C., and Schladow, S. G., 2005. In-flight validation of mid and thermal infrared data from the multispectral thermal imager (MTI) using an automated high altitude validation site at Lake Tahoe CA/NV USA. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 1991–1999.
- Hook, S. J., Vaughan, R. G., Tonooka, H., and Schladow, S. G., 2006. Absolute radiometric in-flight validation of mid and thermal infrared data from ASTER and MODIS using the lake Tahoe CA/NV, USA automated validation site. *IEEE Transactions on Geoscience and Remote Sensing*, **45**(6), 1798–1807.
- Miles, R. H., and Chow, D. H., 1996. Chap. 7. In Razeghi, M. (ed.), *Long Wavelength Infrared Detectors*. Singapore: Gordon and Breach.
- Prabhakara, C., and Dalu, G., 1976. Remote sensing of surface emissivity at 9 μm over the globe. *Journal of Geophysical Research*, **81**(21), 3719–3724.
- Salomonson, V., Barnes, W., Maymon, P., Montgomery, H., and Ostrow, H., 1989. MODIS: advanced facility instrument for studies of the Earth as a system. *IEEE Transactions on Geoscience and Remote Sensing*, **27**, 145–153.
- Tonooka, H., Palluconi, F., Hook, S., and Matsunaga, T., 2005. Vicarious calibration of ASTER thermal infrared bands. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 2733–2746.
- Yamaguchi, Y., Kahle, A. B., Tsu, H., Kawakami, T., and Pniel, M., 1998. Overview of advanced spaceborne thermal emission reflectance radiometer. *IEEE Transactions on Geoscience and Remote Sensing*, **36**, 1062–1071.

Cross-references

[Calibration and Validation](#)
[Calibration, Optical/Infrared Passive Sensors](#)
[Climate Data Records](#)
[Crop Stress](#)
[Data Processing, SAR Sensors](#)
[Emerging Technologies](#)
[Irrigation Management](#)
[Land Surface Emissivity](#)
[Land Surface Temperature](#)
[Observational Platforms, Aircraft, and UAVs](#)
[Observational Systems, Satellite](#)
[Ocean-Atmosphere Water Flux and Evaporation](#)
[Optical/Infrared, Radiative Transfer](#)
[Remote Sensing, Physics and Techniques](#)
[Resource Exploration](#)
[Volcanism](#)
[Water and Energy Cycles](#)
[Water Resources](#)

TRACE GASES, STRATOSPHERE, AND MESOSPHERE

Nathaniel Livesey
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Synonyms

Atmospheric chemistry; Atmospheric composition; Middle atmosphere (common term used to refer collectively to the stratosphere and mesosphere)

Definitions

Trace gases. Minor constituents in Earth's atmosphere typically having abundances from as large as hundreds of parts per million (e.g., CO_2) to parts per trillion (e.g., BrO and OH) or less.

Stratosphere. The region of Earth's atmosphere between ~15 and ~50 km characterized by temperatures that generally increase with increasing altitude.

Mesosphere. The region of Earth's atmosphere between ~50 km and ~80 km characterized by temperatures that generally decrease with increasing altitude.

Introduction

Earth's atmosphere is divided vertically into regions defined by the vertical gradient of temperature. The lowermost region, the troposphere, ranging from the surface to the "tropopause" at 10–15 km altitude, is characterized by temperatures that mainly decrease with increasing altitude. Above is the stratosphere, where temperatures generally increase with height, up to ~50 km where the "stratopause" separates the stratosphere from the mesosphere, within which temperatures generally decrease with altitude again (up to the "mesopause" around 80 km). Collectively, the stratosphere and mesosphere are often referred to as the "middle atmosphere."

Trace gases play disproportionately important roles in these regions, both chemically and radiatively, despite their small abundances ranging from hundreds of parts per million to as low as parts per trillion or less. For example, the absorption of solar ultraviolet radiation by the [stratospheric ozone](#) layer, in addition to shielding terrestrial organisms from harmful exposure, also heats the stratosphere, giving rise to its increase in temperature with increasing altitude. This positive vertical temperature gradient results in vertical atmospheric stability, leading to the stratification for which the region is named. Thermal emission by carbon dioxide (in addition to methane and other minor trace gases) is the main vehicle for radiative cooling in the stratosphere and mesosphere. Despite being the most abundant species in the atmosphere, molecular oxygen and nitrogen play relatively minor roles in the radiative balance and chemistry of these regions.

In addition to trace gases, aerosols also play a significant role in the chemistry of the middle atmosphere. Most notably, key chemical reactions giving rise to the Antarctic ozone hole take place on the surface of polar stratospheric cloud (PSC) particles (see the companion article on stratospheric ozone). In addition, sulfate aerosol is ubiquitous in the lower stratosphere at small background levels, with occasional significant enhancements resulting from injection of SO_2 directly into the stratosphere by large volcanic eruptions (e.g., Mt. Pinatubo in 1991).

In situ observations of the middle atmosphere, while having provided valuable information, are by their nature limited to those available from high-altitude aircraft and balloons, with sounding rockets the only source of in situ

data in the upper stratosphere and mesosphere. Remote sounding has accordingly formed a cornerstone in scientific investigation of the composition of the stratosphere and mesosphere.

Principals and techniques

Information on stratospheric and mesospheric composition can be obtained from observations over a large range of electromagnetic wavelengths, from the microwave (mainly associated with molecular rotational transitions) to the infrared (mainly vibrational transitions) to the visible and ultraviolet (mainly electronic transitions). Signals observed may arise from thermal emission by the atmosphere itself, from atmospheric absorption and/or scattering of solar, lunar, or stellar radiation or of radiation actively emitted, for example, by laser. Observations are made from the ground and from air- or spaceborne platforms. Variations in viewing angle and/or wavelength observed can give information on the vertical distribution of trace gases. Many space-based measurements have been pioneered by balloon or aircraft precursor instruments, and such instruments are frequently used (in conjunction with in situ instrument) for ongoing scientific investigations and to “validate” satellite observations.

Molecular spectral lines are broadened in the atmosphere by a combination of the ensemble of Doppler shifts from the thermal motion of the molecules (“Doppler” broadening) and by collisions with other molecules (“collision” or “pressure” broadening). The latter generally dominates line widths of infrared and microwave signals in the stratosphere, while for visible and ultraviolet wavelengths, Doppler broadening dominates throughout the middle atmosphere. Accurate knowledge of molecular spectroscopy, including line positions, strengths, and broadening parameters, underpins the accuracy of middle atmosphere composition remote sounding, and fundamental laboratory spectroscopy measurements are an integral part of atmospheric composition research programs.

Pressure broadening of spectral lines can provide valuable information on the vertical distribution of trace gases, with frequencies further from line centers conveying information on lower regions of the atmosphere where lines are broad enough to contribute to the observed signals. For wavelengths where pressure broadening is insignificant, vertical distribution information can still be obtained in some viewing geometries by observing in multiple spectral regions having different atmospheric absorptions (and thus penetration depths).

In the stratosphere and lower mesosphere, the atmosphere is generally in local thermodynamic equilibrium (LTE), that is to say, molecular collisions are frequent enough that the kinetic temperature of the air is in equilibrium with the rotational and vibrational temperature of the emitting and absorbing molecules. At lower pressures (typically smaller than 0.1 hPa, ~60 km), where the mean time between collisions increases, some molecular transitions (particularly infrared vibrational transitions) can be

in non-LTE, that is to say, their “vibrational temperature” differs from the underlying atmospheric kinetic temperature. This complicates the interpretation of observed signals.

Most observations give a measure of number density for a particular molecule, either total column or vertical profile. For profile measurements, knowledge of atmospheric density (i.e., temperature and pressure) is needed to convert these to mixing ratio. Typically, this is inferred by also measuring the number density of a well-mixed gas whose abundance is known well (e.g., O₂ or CO₂).

Direct solar absorption and related approaches

Observation of the absorption of solar radiation by the atmosphere provides some of the most precise remotely sensed measurements of stratospheric and mesospheric composition. The strong solar illumination gives a good signal-to-noise ratio for quantification of atmospheric absorption features. Such observations are possible in all spectral regions having significant solar emission, from the infrared to the ultraviolet. Dobson’s ground-based observations of absorption of solar UV by ozone are one of the earliest examples of remote sounding of atmospheric composition. For ground-based solar absorption observations, vertical distribution information can be obtained from measurements at different sun angles (assuming temporal variations are small).

Low-Earth-orbiting satellites can make “occultation” observations of atmospheric absorption of solar radiation during each satellite sunrise and sunset (giving ~30 measurements per day). Observing variations in absorption as the sun rises/sets through the atmosphere gives high vertical resolution (~1 km) information on trace gas distribution. However, horizontal coverage of such measurements is limited. The latitudes of the sunrise and sunset observations drift slowly from day to day, as dictated by the spacecraft orbit, with (in many but not all cases) all latitudes covered in about a month. Increased coverage can be obtained by making lunar and/or stellar occultation observations. However, such measurements have a poorer signal to noise than solar occultation.

There is a rich history of stratospheric and mesospheric composition measurements from solar occultation instruments, with a continuous record available from the 1984 launch of the Stratospheric Aerosol and Gas Experiment II (SAGE II) instrument (McCormick et al., 1989), followed by SAGE III, the Polar Ozone and Aerosol Measurement (POAM) series of instruments (e.g., Lucke et al., 1999), the Halogen Occultation Experiment (HALOE) (Russell et al., 1993) on the Upper Atmosphere Research Satellite (UARS), and, most recently, the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE/FTS) (Bernath et al., 2005), the only such instrument still currently operating. Additional occultation observations have been made by the Atmospheric Trace Molecule Spectroscopy (ATMOS) instrument flown on the Space Shuttle ATLAS program

(Gunson et al., 1996) and the Improved Limb Atmospheric Spectrometer instruments (ILAS I and II) on the Japanese Advanced Earth-Orbiting Satellites (ADEOS I and II) (e.g., Sasano et al., 2001).

Viewing geometries: nadir, limb, and zenith

In addition to occultation, two other geometries have been extensively employed to observe middle atmosphere composition from low-Earth orbit. “Nadir sounding” instruments view the atmosphere directly below the orbiting spacecraft (or in some swath either side of vertical). “Limb sounding” instruments observe the atmosphere “edge on,” tangent to the Earth, scanning their field of view vertically across the atmosphere, offering better vertical resolution (typically 2–5 km) than nadir observations. In addition, the longer path length associated with limb sounding results in stronger signals for trace gases than for nadir sounding. However, this same long path length results in poorer horizontal resolution for limb observations that is, in principle, possible for nadir sounding. For ground-based and airborne observations, zenith (upward looking) observations of scattered sunlight or thermal emission are commonly employed.

UV/visible scattering measurements

Satellite observations of backscattered UV/visible sunlight constitute the longest continuous record of stratospheric composition observations, most notably ozone, from space. The 1978 launch of the Nimbus-7 Total Ozone Mapping Spectrometer (TOMS) and Solar Backscatter Ultraviolet (SBUV) instruments (Heath et al., 1975) heralded the start of nearly continuous observations using this technique. In addition to measuring total ozone column (e.g., TOMS) and [stratospheric ozone](#) profile (e.g., SBUV), column measurements of SO₂ were made from TOMS, along with a measure of aerosol amount (mainly indicating tropospheric aerosols). More recent instruments include the Global Ozone Monitoring Experiment (GOME) on ERS-2 (Burrows et al., 1999), the SCanning Imaging Absorption SpectroMeter for Atmospheric CHartography (SCIAMACHY) instrument on Envisat (Bovensman et al., 1999), the Ozone Monitoring Instrument (OMI) (Levelt et al., 2006) on Aura, and GOME-2 on METOP (Munro et al., 2006). These continue the TOMS record and also measure column abundances of additional species (e.g., NO₂, HCHO). However, these additional gases are mainly of interest in the troposphere (see the related entry on tropospheric composition observations in this volume). By their nature, solar backscatter observations cannot be made at night (months at high latitude during polar night). Most such observations are made from sun-synchronous low-Earth-orbiting spacecraft (where observations are made at fixed local time).

While most UV/visible stratospheric observations have been made in nadir (or swath), a limited number of instruments have observed the scattering of solar UV/visible

(and near-infrared) radiation in a limb geometry. However, the accuracy of such observations is highly dependent on knowledge of spacecraft attitude. UV/visible limb sounding instrument includes SCIAMACHY, OSIRIS on Odin (Llewellyn et al., 2003; Murtagh et al., 2002), and the profiling component of the “Ozone Mapping and Profiling Suite” (OMPS) on the NPOESS (National Polar-orbiting Operational Environmental Satellite System) Preparatory Project (NPP).

Thermal emission observations

In contrast to UV/visible observations, measurements of atmospheric composition from observations of thermal emission in the infrared and microwave can be made both day and night. While nadir sounding at these wavelengths can yield useful information on stratospheric and mesospheric composition (e.g., for ozone), the majority of space-based thermal emission observations of stratospheric and mesospheric composition have been made in a limb-viewing geometry. Pressure broadening of spectral lines in principle enables instruments viewing limb emission to independently measure both composition and atmospheric pressure for a given limb view, making the overall measurement system less sensitive to knowledge of spacecraft attitude than for UV/visible limb sounding.

Limb sounding instruments in the thermal infrared have included the Limb Radiance Inversion Radiometer (LRIR) flown on Nimbus-6, the Nimbus-7 Limb Infrared Monitor of the Stratosphere (LIMS) (Gille and Russell, 1984), Stratospheric and Mesospheric Sounder (SAMS) (Drummond et al., 1980) instruments, the Improved Stratospheric and Mesospheric Sounder (ISAMS) (Taylor et al., 1993) and the Cryogenic Limb Array Etalon Spectrometer (CLAES) on UARS (Roche et al., 1993), the Michelson Infrared Passive Atmospheric Sounder (MIPAS) (Fischer et al., 2008) on Envisat, the Sounding of the Atmosphere using Broadband Emission Radiometry (SABER) instrument (Russell et al., 1999) on the Thermosphere, Ionosphere, Mesosphere Energetics and Dynamics (TIMED) spacecraft, and the High Resolution Dynamics Limb Sounder (HIRDLS) (Gille et al., 2008) on Aura.

Atmospheric composition instruments observing thermal microwave emission in the limb include the Microwave Limb Sounder (MLS) instruments on the UARS and Aura spacecraft (Barath et al., 1993; Waters et al., 2006), the Microwave Atmospheric Sounder (MAS) flown on the space shuttle ATLAS program (Croskey et al., 1992), the Submillimeter Radiometer (SMR) on Odin (Murtagh et al., 2002), and the planned Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES) (Ozeki et al., 2001) to be attached to the International Space Station.

Observations in the microwave region of the spectrum are unaffected by scattering or emission by stratospheric aerosols. By contrast, infrared limb (or occultation) instruments can readily observe stratospheric aerosols

(notably PSCs or strong volcanic enhancements), but the accuracy of their trace gas observations may be reduced by strong aerosol emission and scattering.

Active techniques: LIDAR

While direct measurements of atmospheric absorption between a source and a distant detector are frequently made at ground level, such approaches are impractical in the middle atmosphere (although many in situ instruments work on the same principal but over short distances, often augmented by multipath reflection optics). However, ozone and water vapor in the middle atmosphere can be measured by observing the atmospheric absorption of backscattered laser light from a Light Detection and Ranging (LIDAR) instrument, either ground based or airborne. The backscatter can be either elastic (Rayleigh scattering) or inelastic (Raman scattering). Observations are typically made at multiple wavelengths that have different levels of absorption by the molecule of interest. Observing differences in the backscattered signal received at the different wavelengths gives a measure of the total molecular absorption between the laser source and the atmospheric layer at which the backscatter occurred. In addition, by observing spectral differences, measurements are insensitive to changes in laser output power. High-speed detection of the backscattered signals enables high vertical resolution (typically hundreds of meters) for LIDAR compared to other techniques.

Notable findings from remote sounding observations

Remote sounding observations of trace gases have provided profound insights in to the behavior of the middle atmosphere, in particular, the complex interplay of chemical, dynamical, and radiative processes that govern the budget of [stratospheric ozone](#). For example, the Antarctic “ozone hole” was originally observed in long-term records of ultraviolet absorption over Halley Bay in Antarctica (Farman et al., 1985). Also, ground-based microwave emission sounding at McMurdo Sound confirmed the role of the chlorine monoxide (ClO) radical in the ozone hole (de Zafra et al., 1987), while UARS MLS measurements showed that similar levels of ClO were present in the Arctic stratosphere (Waters et al., 1993).

Remote sounding of the atmospheric distribution of gases with relatively long lifetimes gives valuable insights into transport of air into, out of, and within the stratosphere and mesosphere. Such transport is fundamental to the spatial distribution of stratospheric ozone and also plays an important and complex role in polar chemical ozone loss (see Sect. 6.2 of Solomon (1999) and references therein). Observations of long-lived trace gas distributions are an important means to quantify the slow overturning circulation of air within the middle atmosphere (the so-called Brewer-Dobson Circulation), which may be impacted by changes in climate. Remote sounding observations have also been essential to our understanding

of how air is transported from the troposphere into the stratosphere (e.g., Mote et al., 1995).

While space-based observations have provided valuable global insights into middle atmosphere composition and processes that affect it, the limited lifetime of space-based instruments makes the extraction of long-term trends problematic. The majority of the information on long-term trends and variations in stratospheric and mesospheric composition has come from ground-based observations. Notable among these are those in the “Network for the Detection of Atmospheric Composition Change” (NDACC, formerly the “Network for the Detection of Stratospheric Change,” NDSC). This network brings together a variety of highly accurate measurement techniques, including LIDAR, and passive observations in the ultraviolet, infrared, and microwave (along with in situ sonde observations) at locations across the globe.

Outlook

The number of space-based instruments observing middle atmosphere composition has decreased significantly in recent years. At the time of writing, this overall decreasing trend is set to continue. All of the spacecraft currently making stratospheric and mesospheric observations are now or will shortly be operating beyond their mission design lifetimes. The only confirmed mission to continue observations is the Japanese Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES) instrument, to be attached to the International Space Station (ISS) in 2009. Plans for future middle atmosphere composition observations are unclear, with a long gap anticipated between NASA’s Aura satellite and the expected Global Atmospheric Composition Mission (GACM) and a similar gap likely between ESA’s Envisat and expected Sentinel-5 missions. Although concept missions to fill this gap have been proposed or are in formulation, the only confirmed observations are [stratospheric ozone](#) profile observations by OMPS on NPP.

Conclusion

Our understanding of the chemical, dynamical, and radiative processes that govern the behavior of the stratosphere and mesosphere has advanced dramatically over the last two decades, thanks in large part to the wealth of remote sounding composition observations. Notable findings from remote sounding include the discovery of the ozone hole, insights into its origins (in conjunction with in situ observations), and understanding of the large-scale transport processes that operate in the middle atmosphere.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Barath, F. T., et al., 1993. The upper atmosphere research satellite microwave limb sounder experiment. *Journal of Geophysical Research*, **98**, 10751–10762.
- Bernath, P. D., et al., 2005. Atmospheric Chemistry Experiment (ACE): mission overview. *Geophysical Research Letters*, **32**, L15S01, doi:10.1029/2005GL022386.
- Bovensman, H., Burrows, J. P., Buckwitz, B., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H., 1999. SCIAMACHY: mission objectives and measurement modes. *Journal of the Atmospheric Sciences*, **56**, 127–150.
- Burrows, J. R., et al., 1999. The Global Ozone Monitoring Experiment (GOME): mission concept and first scientific results. *Journal of the Atmospheric Sciences*, **56**, 151–175.
- Croskey, C. L., et al., 1992. The Millimeter Wave Atmospheric Sounder (MAS): a shuttle-based remote sensing experiment. *IEEE Transactions on Microwave Theory and Techniques*, **40**, 1090–1100.
- de Zafra, R. J., Jaramillo, M., Parrish, A., Solomon, P., Connor, B., and Barrett, J., 1987. High concentrations of chlorine monoxide at low latitudes in the Antarctic spring stratosphere: diurnal variation. *Nature*, **328**, 408.
- Drummond, J. R., Houghton, J. T., Peskett, G. D., Rodgers, C. D., Wale, M. J., Whitney, J. G., and Williamson, E. J., 1980. The stratospheric and mesospheric sounder on nimbus 7. *Philosophical Transactions of the Royal Society of London*, **296**, 219–241.
- Farman, J. C., Gardiner, B. G., and Shanklin, J. D., 1985. Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature*, **315**, 207–210.
- Fischer, H., et al., 2008. MIPAS: an instrument for atmospheric and climate research. *Atmospheric Chemistry and Physics*, **8**, 2151–2188.
- Gille, J. C., and Russell, J. M., III, 1984. The limb infrared monitor of the stratosphere: experiment description, performance and results. *Journal of Geophysical Research*, **89**, 5125–5140.
- Gille, J., et al., 2008. High Resolution Dynamics Limb Sounder (HIRDLS): experiment overview, recovery and validation of initial temperature data. *Journal of Geophysical Research*, **113**, D16S43, doi:10.1029/2007JD008824.
- Gunson, M. R., et al., 1996. The Atmospheric Trace Molecule Spectroscopy (ATMOS) experiment: deployment on the ATLAS space shuttle missions. *Geophysical Research Letters*, **23**, 2333–2336.
- Heath, D. F., Krueger, A. J., Roeder, H. A., and Henderson, B. D., 1975. The solar backscatter ultraviolet and total ozone mapping spectrometer (SBUV/TOMS) for nimbus G. *Optical Engineering*, **14**, 323.
- Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Mälikki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H., 2006. The ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1093–1101.
- Llewellyn, E. J., et al., 2003. First results from the OSIRIS instrument on-board Odin. *Sodankylä Geophysical Observatory Publications*, **92**, 1–47.
- Lucke, R. L., et al., 1999. The Polar Ozone and Aerosol Measurement (POAM) III instrument and early validation report. *Journal of Geophysical Research*, **104**, 18785–18799.
- McCormick, M. P., Zawodny, J. M., Velga, R. E., Larsen, J. C., and Wang, P. H., 1989. An overview of SAGE I and II ozone measurements. *Planetary and Space Science*, **37**, 1567–1586.
- Mote, P. W., Rosenlof, K. H., Holton, J. R., Harwood, R. S., and Waters, J. W., 1995. Seasonal variation of water vapor in the tropical lower stratosphere. *Geophysical Research Letters*, **22**, 1093–1096.
- Munro, R., Eisinger, M., Anderson, C., Callies, J., Corpaccioli, E., Lang, R., Lefebvre, A., Livschitz, Y., and Albiñana, A. P., 2006. GOME-2 on MetOp. In *Proceedings of the 2006 EUMETSAT Meteorological Satellite Conference, Helsinki, June 12–16, 2006*, EUMETSAT p. 48.
- Murtagh, D., et al., 2002. An overview of the Odin atmospheric mission. *Canadian Journal of Physics*, **80**, 309–319.
- Ozeki, H., et al., 2001. Development of superconducting submillimeter-wave limb emission sounder (JEM/SMILES) aboard the International Space Station. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 4540, pp. 209–220, SPIE.
- Roche, A. E., Kumer, J. B., Mergenthaler, J. L., Ely, H. A., Uplinger, W. H., Potter, J. F., James, T. C., and Sterritt, L. W., 1993. The Cryogenic Limb Array Etalon Spectrometer (CLAES) on UARS: experiment description and performance. *Journal of Geophysical Research*, **98**, 10763–10775.
- Russell, J. M., III, et al., 1993. The halogen occultation experiment. *Journal of Geophysical Research*, **93**, 10777–10798.
- Russell III, J. M., Mlynczak, M. G., Gordley, L. L., Tansock, J., and Esplin, R.: An overview of the SABER experiment and preliminary calibration results. *Proc. SPIE*, **3756**, doi:10.1117/12.366382, 277–288, 1999.
- Sasano, Y., Yokota, T., Nakajima, H., Sugita, T., and Kanzawa, H., 2001. ILAS-II instrument and data processing system for stratospheric ozone layer monitoring. *Proceedings of SPIE*, **4150**, 106–114.
- Solomon, S., 1999. Stratospheric ozone depletion: a review of concepts and history. *Reviews of Geophysics*, **37**, 275–316.
- Taylor, F. W., et al., 1993. Remote sensing of atmospheric structure and composition by pressure modulation radiometry from space: the ISAMS experiment on UARS. *Journal of Geophysical Research*, **98**, 10799–10814.
- Waters, J. W., Froidevaux, L., Read, W. G., Manney, G. L., Elson, L. S., Flower, D. F., Jarnot, R. F., and Harwood, R. S., 1993. Stratospheric ClO and ozone from the microwave limb sounder on the upper atmosphere research satellite. *Nature*, **362**, 597–602.
- Waters, J. W., et al., 2006. The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1075–1092.

Cross-references

Stratospheric Ozone

TRACE GASES, TROPOSPHERE - DETECTION FROM SPACE

Pietermel F. Levelt^{1,2}, J. P. Veefkind^{1,3} and K. F. Boersma^{1,4}

¹Koninklijk Nederlands Meteorologisch Instituut (KNMI), De Bilt, The Netherlands

²Delft University of Technology, Eindhoven, The Netherlands

³Eindhoven University of Technology, Eindhoven, The Netherlands

⁴Technical University Eindhoven (TUE), Eindhoven, The Netherlands

Introduction

The increase of human activity over the last 100 years has resulted in an enormous increase in atmospheric

concentrations of various gases of anthropogenic origin, polluting the atmosphere and changing climate on Earth. The detection of the [stratospheric ozone](#) hole raised awareness that human activities can lead to dramatic changes in atmospheric composition, with large consequences for life on Earth. The satellite maps of air pollution and greenhouse gases in the troposphere (this is the lower 10–16 km of the atmosphere) over the last 5–10 years made people realize that our lifestyle does not only influence the ozone layer but also our daily living environment and the air we breathe.

Trace gases in the troposphere

The atmosphere consists of nitrogen (78 %) and oxygen (21 %), with only a minor contribution from other gases (1 %), which are called trace gases due to their low abundance. Although their abundance is low, trace gases play a key role in climate change and air quality. Ozone (O_3) is one of the most abundant minor trace gases. Ozone resides mainly in the stratosphere – the so-called ozone layer – protecting us from the harmful ultraviolet radiation from the Sun because it efficiently absorbs this high-energy radiation. Ozone also plays an important role in the troposphere, where it is a greenhouse gas as well as an air pollutant. Other important tropospheric trace gases are carbon dioxide (CO_2), water vapor (H_2O), methane (CH_4), nitrogen oxides (NO_x), sulfur oxides (SO_x), carbon monoxide (CO), NH_3 and VOCs (volatile organic compounds) like formaldehyde ($HCHO$). CO_2 and H_2O are important greenhouse gases. H_2O plays a major role in the hydrological cycle and is the gas that in reaction with O_3 forms the highly reactive radical OH . OH is the key species in cleansing the atmosphere of pollutants. All other mentioned gases play a role as greenhouse gas or air pollutant and/or as a precursor for a greenhouse gas, aerosols, or air pollutant. For example, NO_2 is an air pollutant and an important precursor for the greenhouse gas ozone and aerosols, as the air pollutant SO_2 is also an important precursor for aerosols. Other nongaseous atmospheric constituents that are of paramount importance for climate change and air pollution are aerosols and clouds.

The trace gases CO_2 , CH_4 , NO_2 , SO_2 , $HCHO$, NH_3 and CO , which can be observed from space and are increasing due to human activities, will therefore be the focus of this entry. H_2O is a natural gas and will therefore not be considered. OH in the troposphere is undetectable from space.

History of detection of trace gases from space

In the early 1970s, the first measurements of (stratospheric) ozone were made from space by the SBUV satellite instrument (Hilsenrath et al., [1995](#)), later followed by the TOMS instrument (McPeters et al., [1996](#)). These were the first measurements of the chemical composition of the atmosphere from space. In addition to stratospheric ozone, TOMS also observed SO_2 from volcanic eruptions. GOES was used to retrieve aerosol optical depth (Fraser et al., [1984](#)). Fishman et al. ([1987](#)) used ozone columns from

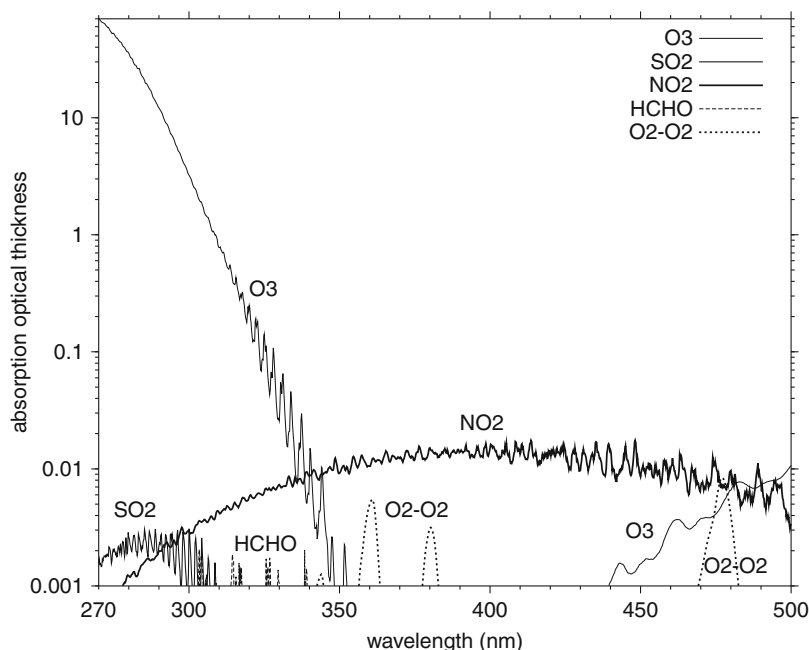
TOMS to investigate an episode with ozone smog over the eastern United States. An important step forward was made with the launch of the GOME instrument in 1995 (Burrows et al., [1999](#)), which could measure several of the minor trace gases using the backscatter technique. This instrument was followed by SCIAMACHY (2002, Bovensmann et al., [1999](#)), OMI (2004, Levelt et al., [2006](#)), and GOME-2 (2006, Callies et al., [2000](#)). Apart from the solar backscatter technique, measurements of tropospheric trace gases can also be made in the thermal infrared. MOPITT (1999, Deeter, [2009](#)), AIRS (2002, Aumann et al., [2003](#)), TES (2004, Beer, [2006](#)), and IASI (2007, Blumstein et al., [2004](#)) are instruments that use this wavelength range.

All instruments that have been observing tropospheric trace gases were on board satellites in a Sun-synchronous orbit, providing at most one (backscatter techniques) or two (thermal infrared) measurements at the same local time per day.

Retrievals and validation

What can be detected from space on tropospheric trace gases? Since the species of interest reside in the troposphere, and the instruments are orbiting the Earth at around 500–800 km altitude, observing in nadir, it is quite a challenge to detect the minor constituents in the lower 10–15 km of the atmosphere from space. There are basically two methods used to measure the troposphere, both based on passive remote sensing techniques: the solar backscatter technique and the thermal infrared technique. In the solar backscatter technique, the instrument measures the solar radiation that has been absorbed and scattered by the atmosphere. This so-called Earth radiance spectrum contains the specific absorption features of the molecules of interest (due to electronic-vibrational-rotational transitions in the molecule, see [Figure 1](#)). In the thermal infrared technique, the thermal emission of the Earth-atmosphere system is measured, revealing the specific absorption features of the trace gases (see [Figure 2](#)). Due to the specific spectroscopic “fingerprint” of the molecule, it is possible to determine the trace gas concentration, using very sensitive detectors and optimally designed optical instruments. The solar backscatter instruments usually provide tropospheric columns of the trace gases. The technique has the advantage to be sensitive to the surface, since the atmosphere is transparent in the visible wavelength range. With the thermal infrared technique, also some vertical information can be obtained, approximately two layers in the troposphere, but the sensitivity to the surface is less so that accurate total column amounts are more challenging.

The retrieval of tropospheric trace gas concentrations from satellite measurements is a so-called ill-posed problem. For ill-posed problems, the information from the measurement is not enough to derive all unknowns independently. This means that to retrieve trace gas concentrations from satellite measurements, assumptions have to be



Trace Gases, Troposphere - Detection from Space, Figure 1 Total absorption optical thickness for tropospheric trace gases in the UV-vis wavelength range, as calculated for a mid-latitude atmosphere in summer. Note that a logarithmic scale is used, thus the absorption by ozone decreases by six orders of magnitude in the wavelength range from 270 to 360 nm (Image courtesy: P. Stammes (KNMI)).

made on, for example, the reflectivity of the surface and the approximate vertical distribution of the trace gas. One of the most widely used retrieval techniques is optimal estimation (Rodgers, 2000). For the solar backscatter technique, also differential optical absorption spectroscopy (DOAS) retrieval is widely used (Platt, 1994).

Satellite measurements of tropospheric trace gases are hampered by clouds that reflect radiation to space, effectively screening the gas concentrations in the most polluted layer below the clouds. For fully clouded scenes algorithms are therefore fundamentally limited to retrieving above-cloud concentrations. But in situations with partial cloud cover, radiance signals still contain information on trace gas concentrations in the lowest layers, and tropospheric column retrievals are possible. By reducing the size of the ground pixel, the probability of encountering a completely cloud-free pixel increases. Therefore, recent developments have focused on limiting the pixel size, thereby increasing the amount of measurement samples in order to obtain more cloud-free observations while retaining global coverage.

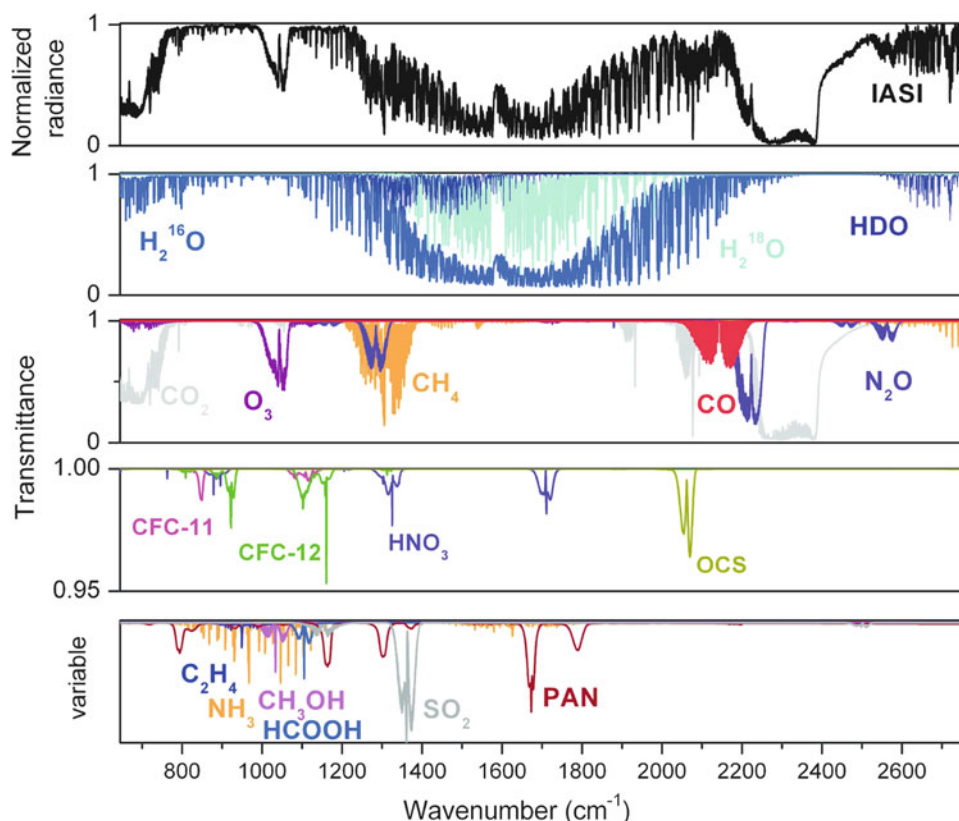
In order to check their accuracy, satellite measurements are validated by comparisons with ground-based and aircraft measurements and atmospheric models. Ground-based and aircraft campaigns are often performed for validation for specific atmospheric conditions and are thus limited in scope. Networks of ground-based instruments of known quality are thus a valuable tool for validation of satellite measurements over the whole mission lifetime.

Extensive networks exist only for a limited set of trace gases (for instance, ozone and CO_2). For other trace gases, like NO_2 , SO_2 , and HCHO , the development of consistent ground-based networks is much called for, but only starting.

The advantages of satellite observations

The capability of satellite instruments to measure the tropospheric pollution first became apparent by measurements of GOME on NO_2 (Leue et al., 2001) and MOPITT on CO (Edwards et al., 2004). These instruments enabled for the first time global measurements of pollutants in the troposphere in the form of maps of monthly or yearly averaged concentrations. Up to that point, only models provided such information, and independent data to validate these models was sparse. SCIAMACHY extended the measurements of the troposphere to greenhouse gases like methane (Frankenberg et al., 2005). The GOSAT satellite (Hamazaki et al., 2004) was launched in 2009 is dedicated to measure CH_4 and CO_2 with very high accuracy. Recent results suggest that GOSAT CO_2 retrievals are useful to constrain surface fluxes of CO_2 (Butz et al., 2011). In 2014 NASA's OCO satellite (Crisp et al., 2004) will be launched dedicated to measure CO_2 .

Because of its small pixels and daily global coverage, OMI provides much more measurements of the troposphere than previous instruments were able to. Instead of monthly averaged maps, OMI is able to provide daily maps. In Figure 3, four consecutive frames are shown of



Trace Gases, Troposphere - Detection from Space, Figure 2 IASI-normalized radiance spectrum (*top panel*, recorded by IASI/ MetOp, over west of Australia, on December 20, 2006) and transmittance simulated molecular contributions to identify the main absorbing gases (*middle panels*) and the weaker (*bottom panel*) (From Clerbaux et al., 2009).

OMI measurements. The figure clearly shows the Sunday dip in NO_2 concentration due to reduction of traffic during the weekend resulting in less NO_2 emissions.

These instruments clearly showed the unique capability of satellite measurements to obtain global coverage and consistent quality of the measurements. For instance, ground-based networks for most of the tropospheric trace gases are sparse and the quality of the data is sometimes station dependent.

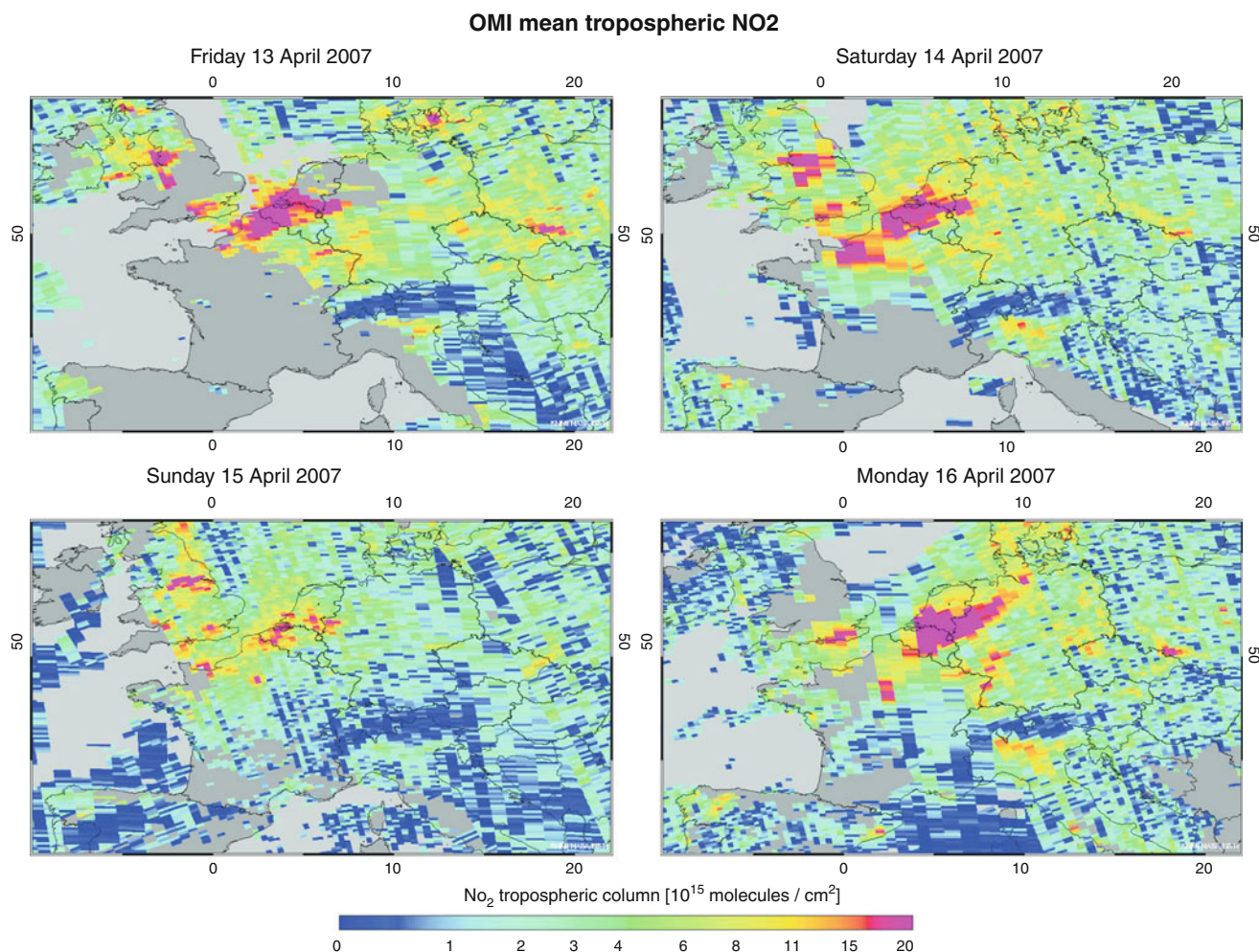
The chemistry of the troposphere is complex and involves many trace gases and chemical reactions. Combining the observations of trace gases from satellite instruments is therefore a great advantage. The added value of satellite observations of key species is illustrated in Figure 4. This figure shows the prominent global impact of human activities on the composition of the troposphere. Figure 4 shows satellite maps of concentrations of NO_2 , CO, CH_4 , HCHO, and particulate matter. Also shown is a map of the population density. The sources for NO_2 and CO are fossil fuel combustion by power plants, industry, and traffic. For CH_4 , important sources are wetlands (including rice agriculture), energy, livestock, landfills and waste, and biomass burning. Formaldehyde sources include biogenic emissions, as

well as biomass burning and fossil fuel burning. Most of the particulate matter is formed in the atmosphere from the trace gases.

Research themes and operational applications

Measurements of tropospheric trace gases are important for air quality monitoring and forecasting. Especially, observations of longer-lived trace gases such as CO have proven useful in analyzing the impact of distant sources to local air pollution levels (e.g., Gloudemans et al., 2009). Air quality prediction systems increasingly use satellite observations to improve their forecasting capability.

Satellites provide top-down constraints on emission inventories, which traditionally rely on rapidly outdated bottom-up estimates, and generally go unchecked by measurements. As a result of the unique global character of satellite data and their consistency (one retrieval algorithm for all measurements), satellite measurements provide an exceptional tool for checking emission databases. Examples include Lamsal et al. (2011) who combined a top-down approach with a bottom-up a priori inventory to produce an optimal a posteriori inventory for NO_x emissions. Mijling et al. (2012) recently applied a new



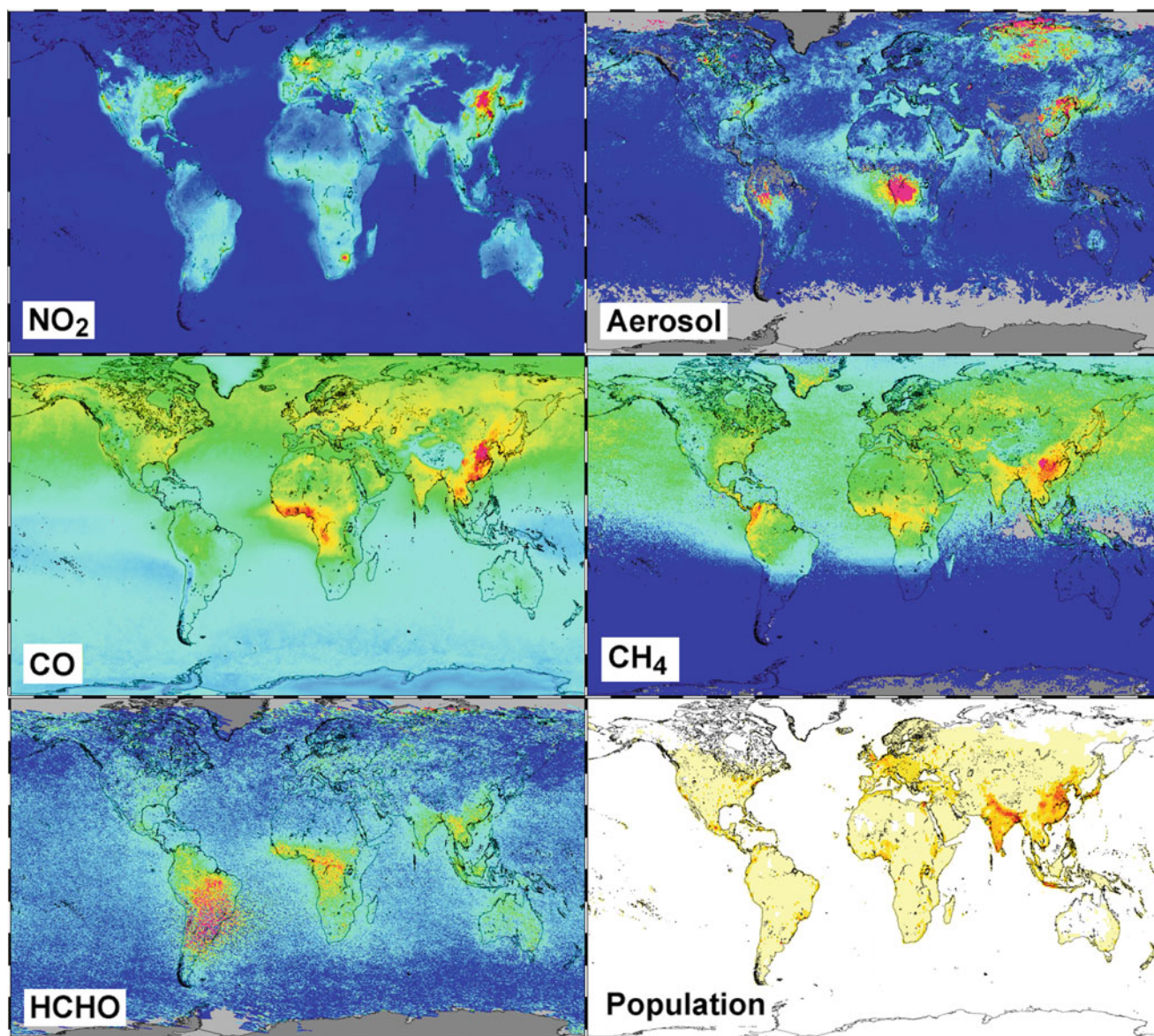
Trace Gases, Troposphere - Detection from Space, Figure 3 Four consecutive frames of OMI tropospheric NO₂ measurements. The figure clearly shows the Sunday dip in NO₂ concentration due to reduction of traffic resulting in less NO₂ emissions (Image courtesy: R. Noordhoek (KNMI); data courtesy: KNMI/NASA/NSO and www.temis.nl).

inversion technique to use daily satellite observations for fast updates of emission estimates at high spatial resolution ($25 \times 25 \text{ km}^2$) for eastern Asia. VOC emissions were first derived from GOME retrievals of HCHO over North America using an inversion based on the relationship between HCHO columns and VOC emissions scaled by their HCHO yields (Palmer et al., 2003). Important developments include simultaneous multispecies inversions to infer CO and NO_x emissions from MOPITT and GOME information (Müller and Stavrakou, 2005), and recently, Beirle et al. (2011) estimated megacity NO_x emissions and lifetimes directly from the decay-with-distance of NO₂ concentrations downwind of megacities.

The satellite data can also be used to evaluate models. Comparison of air quality models – traditionally focusing on the lower troposphere – to satellite measurements, for instance, has shown the need to accurately represent the free troposphere in these models (Blond et al., 2007).

Using SCIAMACHY NO₂ columns, Bertram et al. (2005) were able to improve the model description of microbial soil NO_x emissions that depend on humidity, fertilizer application, and temperature.

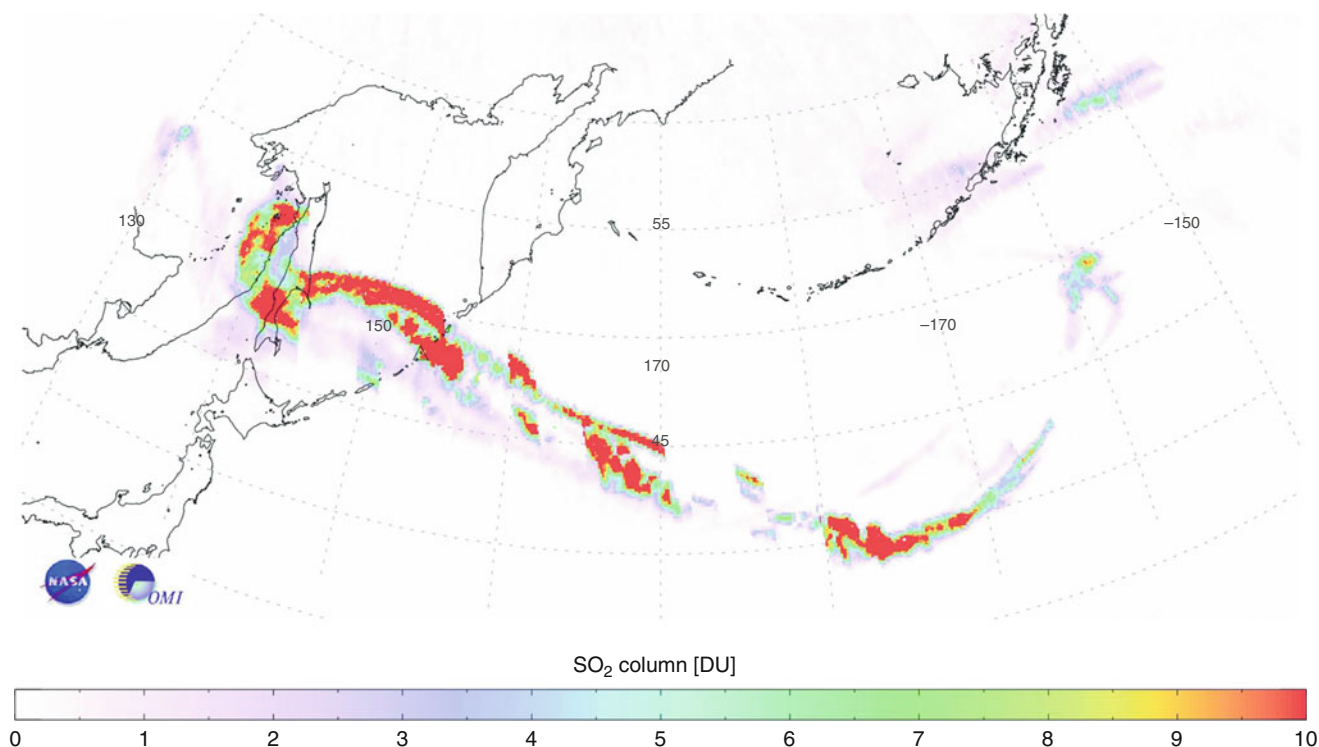
Satellite data are also very suitable to extent our knowledge on the atmospheric composition by analyzing the long-term records. Richter et al. (2005) showed the increase of NO₂ pollution over eastern Asia over the last 10 years. Van der A et al. (2008) extended this work and analyzed spatially resolved trends in NO₂ concentrations for the entire globe. Recently, Castellanos and Boersma (2012) showed that reductions in NO₂ concentrations over Europe reflect a combination of environmental policy and economic activity. Especially the 2008–2009 global economic recession triggered a distinct change in anthropogenic activity in Europe, indicated by sharp downturns in grossdomestic product and industrial production, and, consequently, in air pollution.



Trace Gases, Troposphere - Detection from Space, Figure 4 Global concentration maps. Color scale concentrations range from red, via yellow and green, to blue, which represents very high, high, medium, and low values, respectively. For population (lower right) white represents small, and dark brown, red represents large population numbers. From top-left to bottom-right: (a) OMI tropospheric column NO_2 , average for 2007, (b) POLDER/PARASOL (Tanré et al., 2011) fine-mode aerosol optical thickness, June–August 2006, (c) MOPITT CO average mixing ratio between 0–2 km and 7 km altitude, 7 year average (2000–2007), (d) SCIAMACHY methane column-averaged mixing ratio, Jan 2003 – Oct 2005, (e) SCIAMACHY formaldehyde, average for year 2004, (f) World population map from the Center for International Earth Science Information Network (CIESIN), Columbia University, and Centro Internacional de Agricultura Tropical. (Image courtesy Henk Eskes, KNMI and http://esamultimedia.esa.int/docs/SP1313-6_TRAQ.pdf).

For air quality prediction, usually data assimilation systems are used, where satellite data and ground-based data are integrated with the model data in order to improve the forecast. This technique is only effective when the data are available within a few hours of measurement, also called near real time (NRT). Nowadays, measurements from most backscatter instruments can be obtained within this time frame (e.g., Boersma

et al., 2007). In the thermal infrared, the IASI instrument is also delivering in NRT. The NRT SO_2 measurements from volcanoes are used for aviation control, relaying aircraft after a volcanic outburst. Figure 5 shows a SO_2 plume from a volcanic outburst as observed by OMI in June 2009. The observations clearly indicate the locations that should be avoided by intercontinental flights.



Trace Gases, Troposphere - Detection from Space, Figure 5 The image shows average column sulfur dioxide concentrations between 11 and 17 June 2009, based on data from OMI. The high SO₂ amounts originate from an eruption of the Sarychev Volcano (Kuril Islands, northeast of Japan) on June 12, 2009 (Images like these are used to redirect aircrafts. Image courtesy: S. Carn (Michigan Technology University)).

Future

To improve the measurements of the tropospheric trace gases, smaller ground pixels are essential to reduce cloud contamination and obtain highly resolved information on suburban scales. New instruments will target at a $5 \times 5 \text{ km}^2$ spatial resolution. In 2015 the TROPOMI instrument will be launched on ESA's sentinel-5 precursor (Veefkind et al., 2012) with a spatial resolution of $7 \times 7 \text{ km}^2$. From 2020 onward this instrument will be followed by ESA's sentinel 5 instrument on METOP-SG as part of the EU Copernicus program (Ingmann et al., 2012). Special cloud-observing capabilities are in development so that extra cloud information can be measured and the accuracy of the trace gas retrievals for partly cloudy pixels can be improved. By developing combined retrieval techniques that take advantage of backscatter as well as the thermal infrared spectral measurements, vertical profiles of tropospheric trace gases with more than two pieces of information in the troposphere are anticipated. Such combined techniques can be used when both instruments are located on the same satellite. Adding a dedicated aerosol instrument to such a platform would allow for simultaneous observations of all anthropogenic influences on the troposphere for the first time. The METOP-SG satellite is expected to have this combination of instruments on board (Berger et al., 2012).

Apart from a Sun-synchronous orbit, also non-Sun-synchronous as well as geostationary orbits could be used. The Sun-synchronous orbit will provide full global measurements with one or two observations a day. From the geostationary orbit, hourly observations can be made for a relatively small part of the world (e.g., Europe or North America). The non-Sun-synchronous-inclined orbit (see e.g., http://esamultimedia.esa.int/docs/SP1313-6_TRAQ.pdf) combines global coverage (apart from the polar regions) and frequent measurements (about five per day). Two to three of these non-Sun-synchronous satellites are needed to provide high temporal measurements throughout the year for the whole globe, except the poles. Which orbit to use is dependent on the specific scientific or operational purposes of the satellite. For climate observations, the global coverage of the measurements is of paramount importance. For regional air quality, the high temporal sampling is more important than global coverage. In the timeframe 2017-2020 three geostationary missions are planned: the Sentinel 4 (Ingmann et al., 2012) as part of the EU Copernicus programme over Europe, the South Korean GEMS (Kim et al., 2012) over Southeast Asia, and the U.S. TEMPO (Fishman et al., 2012) over North America. These instrument form a constellation of geostationary air quality missions, can be linked to the Low-Earth-Missions like the sentinel 5 precursor and

METOP-SG that provide the global coverage and thus interconnect the regional missions.

Summary

In the last 15 years, observing trace gases in the troposphere from space has been a tremendously fast developing field. Sophisticated retrieval techniques have been developed in order to obtain tropospheric trace gas concentrations, and these techniques have been validated successfully. Satellite instruments now provide daily observations of the chemical composition of the troposphere with spatial detail as fine as urban scale. Satellite observations have been successfully used to provide constraints on emissions of pollutants, to evaluate and improve atmospheric models, and to identify trends in air pollution. In the future, pixel sizes will even decrease further, and combined measurements from backscatter and thermal infrared will provide better vertical profiles in the troposphere.

Bibliography

- Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L. M., Revercomb, H., Rosenkranz, P. W., Smith, W. L., Staelin, D. H., Strow, L. L., and Susskind, J., 2003. AIRS/AMSU/HSB on the aqua mission: design, science objectives, data products, and processing systems. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(2), 253–264, doi:10.1109/TGRS.2002.808356.
- Beer, R., 2006. TES on the aura mission: scientific objectives, measurements, and analysis overview. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1102–1105.
- Beirle, S., Boersma, K. F., Platt, U., Lawrence, M. G., and Wagner, T., 2011. Megacity emissions and lifetimes of nitrogen oxides probed from space. *Science*, **333**, 1737–1739, doi:10.1126/science.1207824, 2011.
- Berger, M., Moreno, J., Johannessen, J. A., Levelt, P. F., and Hanssen, R. F., 2012. ESA's sentinel missions in support of Earth system science. *Remote Sensing of Environment*, **120**, 84–90, doi:10.1016/j.rse.2011.07.023.
- Bertram, T. H., Heckel, A., Richter, A., Burrows, J. P., and Cohen, R. C., 2005. Satellite measurements of daily variations in soil NO_x emissions. *Geophysical Research Letters*, **32**, L24812, doi:10.1029/2005GL024640.
- Blond, N., Boersma, K. F., Eskes, H. J., van der A, R. J., Van Roozendaal, M., De Smedt, I., Bergametti, G., and Vautard, R., 2007. Intercomparison of SCIAMACHY nitrogen dioxide observations, in situ measurements and air quality modeling results over Western Europe. *Journal of Geophysical Research*, **112**, doi:10.1029/2006JD007277.
- Blumstein, D., Chalon, G., Carlier, T., Buil, C., Hébert, P., Maciaszek, T., Ponce, G., Phulpin, T., Tournier, B., Siméoni, D., Astruc, P., and Clauss, A., 2004. IASI instrument: technical overview and measured performances. In *Proceedings SPIE Conference*, Denver, Aug 2004, vol. 5543, pp. 196–207 doi: 10.1117/12.560907.
- Boersma, K. F., Eskes, H. J., Veeffkind, J. P., Brinksma, E. J., van der A, R. J., Sneep, M., van den Oord, G. H. J., Levelt, P. F., Stammes, P., Gleason, J. F., and Bucsela, E. J., 2007. Near-real time retrieval of tropospheric NO₂ from OMI. *Atmospheric Chemistry and Physics*, **7**, 2103–2118.
- Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H., 1999. SCIAMACHY: mission objectives and measurement modes. *Journal of Atmospheric Sciences*, **56**, 127–155.
- Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V. V., Ladstätter-Weissenmayer, A., Richter, A., De Beek, R., Hoogen, R., Bramstedt, K., Eichmann, K. W., Eisinger, M., and Perner, D., 1999. The Global Ozone Monitoring Experiment (GOME): mission concept and first scientific results. *Journal of Atmospheric Sciences*, **56**, 151–175.
- Butz, A., Guerlet, S., Hasekamp, O., Schepers, D., Galli, A., Aben, I., Frankenberg, C., Hartmann, J.-M., Tran, H., Kuze, A., Keppel-Aleks, G., Toon, G., Wunch, D., Wennberg, P., Deutscher, N., Griffith, D., Macatangay, R., Messerschmidt, J., Notholt, J., and Warneke, T., 2011. Toward accurate CO₂ and CH₄ observations from GOSAT. *Geophysical Research Letters*, **38**(14), L14812, doi:10.1029/2011GL047888.
- Callies, J., Corpaccioli, E., Eisinger, M., Hahne, A., and Lefebvre, A., 2000. GOME-2 – MetOp's second generation sensor for operational ozone monitoring. *ESA Bulletin*, **102**, 28–36.
- Castellanos, P., and Boersma, K. F., 2012. Reductions in nitrogen oxides over Europe driven by environmental policy and economic recession. *Scientific Reports*, **2**(2), 265, doi:10.1038/srep00265.
- Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F., 2009. Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder. *Atmospheric Chemistry and Physics*, **9**, 6041–6054, doi:10.5194/acp-9-6041-2009.
- Crisp, D., Atlas, R. M., Breon, F.-M., Brown, L. R., Burrows, J. P., Ciaia, P., Connor, B. J., Doney, S. C., Fung, I. Y., Jacob, D. J., Miller, C. E., O'Brien, D., Pawson, S., Randerson, J. T., Rayner, P., Salawitch, R. J., Sander, S. P., Sen, B., Stephens, G. L., Tans, P. P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., Kuang, Z., Chudasama, B., Sprague, G., Weiss, B., Pollock, R., Kenyon, D., and Schroll, S., 2004. The Orbiting Carbon Observatory (OCO) mission. *Advances in Space Research*, **34**(4), 700–709.
- Deeter, M. N., 2009. MOPITT (Measurements of Pollution in the Troposphere) Validated Version 4 Product User's Guide. Available from http://www.acd.ucar.edu/mopitt/v4_users_guide_val.pdf.
- Edwards, D. P., Emmons, L. K., Hauglustaine, D. A., Chu, D. A., Gille, J. C., Kaufman, Y. J., Pétron, G., Yurganov, L. N., Giglio, L., Deeter, M. N., Yudin, V., Ziskin, D. C., Warner, J., Lamarque, J.-F., Francis, G. L., Ho, S. P., Mao, D., Chen, J., Grechko, E. I., and Drummond, J. R., 2004. Observations of carbon monoxide and aerosols from the Terra satellite: northern hemisphere variability. *Journal of Geophysical Research*, **109**, D24202, doi:10.1029/2004JD004727.
- Fishman, J., Vukovich, F. M., Cahoon, D., and Shipman, M. C., 1987. The characterization of an air pollution episode using satellite total ozone measurements. *Journal of Applied Meteorology*, **26**, 1638–1654.
- Fishman, J., and co-authors, 2012. The United STATES' next generation of atmospheric composition and coastal ecosystem measurements: NASA's geostationary coastal and Air pollution events (GEO-CAPE), mission. *Bulletin of the American Meteorological Society*, **93**, 1547–1566. <http://dx.doi.org/10.1175/BAMS-D-11-00201.1>.
- Fraser, R. S., Kaufman, Y. J., and Mahoney, R. L., 1984. Satellite measurements of aerosol mass transport. *Atmospheric Environment*, **18**, 2577–2584.
- Gloudemans, A. M. S., de Laat, A. T. J., Schrijver, H., Aben, I., Meirink, J. F., and van der Werf, G. R., 2009. SCIAMACHY CO over land and oceans: 2003–2007 inter annual variability. *Atmospheric Chemistry and Physics*, **9**, 3799–3813, doi:10.5194/acp-9-3799-2009.
- Hamazaki, T., Kuze, A., Kondo, K., 2004. Sensor system for Greenhouse Gas Observing Satellite (GOSAT). In *Proceedings SPIE*

- Conference, Denver, Aug 2004, vol. 5543, pp. 275–282 doi: 10.1117/12.560589.
- Hilsenrath, E., Cebula, R. P., Deland, M. T., Laamann, K., Taylor, S., Wellemeyer, C., and Bhartia, P. K., 1995. Calibration of the NOAA-11 Solar Backscatter Ultraviolet (SBUV/2) ozone data set from 1989 to 1993 using in-flight calibration data and SSBUV. *Journal of Geophysical Research*, **100**, 1351–1366.
- Ingmann, P., Veihelman, B., Langen, J., Lamarre, D., Stark, H., and Courreges-Lacoste, G. B., 2012. Requirements for the GMES Atmospheric Service and ESAs implementation concept: Sentinels-4/-5 and -5p. *Remote Sensing of the Environment*, **120**, 58–69, doi:10.1016/j.rse.2012.01.023.
- Kim, J., 2012. GEMS(Geostationary Environment Monitoring Spectrometer) onboard the GeoKOMPSAT to Monitor Air Quality in high Temporal and Spatial Resolution over Asia-Pacific Region, *EGU General Assembly 2012*, held 22–27 April, 2012 in Vienna, Austria., p. 4051, EGU2012–4051.
- Lamsal, L. N., Martin, R. V., Padmanabhan, A., van Donkelaar, A., Zhang, Q., Sioris, C. E., Chance, K., Kurosu, T. P., and Newchurch, M. J., 2011. Application of satellite observations for timely updates to global anthropogenic NO_x emission inventories. *Geophysics Research and Letters*, **38**, L05810, doi:10.1029/2010GL046476.
- Leue, C., Wenig, M., Wagner, T., Klimm, O., Platt, U., and Jähne, B., 2001. Quantitative analysis of NO_x emissions from global ozone monitoring experiment satellite image sequences. *Journal of Geophysical Research*, **106**, 5493–5505.
- Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Mälkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J., and Saari, H., 2006. The ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(5), 1093–1101, doi:10.1109/TGRS.2006.872333.
- McPeters, R. D., Bhartia, P. K., Krueger, A. J., Herman, J. R., Schlesinger, B. M., Wellemeyer, C. G., Seftor, C. J., Jaross, G., Taylor, S. L., Swissler, T., Torres, O., Labow, G., Byerly, W., Cebula, R., McPeters, R. D., Bhartia, P. K., Krueger, A. J., Herman, J. R., Schlesinger, B. M., Wellemeyer, C. G., Seftor, C. J., Jaross, G., Taylor, S. L., Swissler, T., Torres, O., Labow, G., Byerly, W., and Cebula, R. P., 1996. *Nimbus-7 Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide*. Washington, DC: NASA. NASA Reference Publication, Vol. 1384.
- Mijling, B., and Van der A, R. J., 2012. Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale. *Journal of Geophysics and Research*, **117**, D17302, doi:10.1029/2012JD017817.
- Müller, J.-F., and Stavrou, T., 2005. Inversion of CO and NO_x emissions using the adjoint of the IMAGES model. *Atmospheric Chemistry and Physics*, **5**, 1157–1186.
- Palmer, P. I., Jacob, D. J., Fiore, A. M., Martin, R. V., Chance, K., and Kurosu, T. P., 2003. Mapping isoprene emissions over North America using satellite observations of formaldehyde columns. *Journal of Geophysical Research*, **108**, 4180, doi:10.1029/2005JD002153.
- Platt, U., 1994. Differential optical absorption spectroscopy (DOAS). In Sigrist, M. W. (ed.), *Air Monitoring by Spectroscopic Techniques*. New York: Wiley, pp. 27–84.
- Richter, A., Burrows, J. P., Nüß, H., Claire Granier, C., and Niemeier, U., 2005. Increase in tropospheric nitrogen dioxide over China observed from space. *Nature*, **437**, 129–132.
- Rodgers, C. D., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. River Edge: World Science.
- Tanré, O., Bréon, F. M., Deuzé, J. L., Dubovik, O., Ducos, F., François, P., Goloub, P., Herman, M., Lifermann, A., Waquet, F., 2011. Remote sensing of aerosols by using polarized, directional and spectral measurements within the A-Train: the PARASOL mission. *Atmospheric Measurement Techniques*, **4**, 1383–1395. doi:10.5194/amt-4-1383-1395-2011.
- van der A, R. J., Eskes, H. J., Boersma, K. F., van Noije, T. P. C., Van Roozendaal, M., De Smedt, I., Peters, D. H. M. U., Kuenen, J. J. P., and Meijer, E. W., 2008. Identification of NO₂ sources and their trends from space using seasonal variability analyses. *Journal of Geophysical Research*, **113**, D04302, doi:10.1029/2007JD009021.
- Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F., 2012. TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote Sensing of Environment*, **120**, 70–83, doi:10.1016/j.rse.2011.09.027.

Cross-references

[Aerosols](#)
[Air Pollution](#)
[Climate Monitoring and Prediction](#)
[Data Assimilation](#)
[Geophysical Retrieval, Inverse Problems in Remote Sensing](#)
[Optical/Infrared, Atmospheric Absorption/Transmission, and Media Spectral Properties](#)
[Stratospheric Ozone](#)
[Trace Gases, Stratosphere, and Mesosphere](#)
[Ultraviolet Remote Sensing](#)
[Ultraviolet Sensors](#)

TRAFFICABILITY OF DESERT TERRAINS

Charles Hibbitts
 Applied Physics Laboratory, Laurel, MD, USA

Definition

Trafficability (of desert terrains): The ability to traverse a surface in a cart vehicle.

The trafficability of desert terrain is controlled by soil strength, surface roughness, and the propensity for a surface to generate dust (dust loading) that would obscure vision or impair engine performance. This entry focuses on the most important and treacherous factor in desert trafficability, soil strength. Surface roughness and dust loading are mentioned in context with specific terrain types. Finally, methods for remotely estimating these factors are briefly discussed. Perhaps the most difficult to recognize, and likely the most important factor for trafficability in desert terrains, is soil strength, which itself is dependent upon several factors that can be measured or estimated: soil moisture, grain (or clast) size, and composition. It is the combination of soil strength and vehicle footprint (the area over which the weight of the vehicle is distributed) that dominates a surface's trafficability.

Contributing factors to soil strength

Soil grain size

In general, as the amount of fine-grained material ($<125\ \mu\text{m}$, or 200 sieve) in a soil increases, its load-bearing strength decreases and trafficability worsens. The “critical depth,” defined as the depth at which soils provide sufficient strength to support a specific vehicle type, increases, and the terrain becomes more difficult to traverse as the fine-grained component increases. Coarse soils under both wheeled and tracked vehicles generally have a critical depth of no more than 6" (wheels sink $\sim 6''$ until traction is obtained), while the critical depth of support for wheeled vehicle on fine-grained soils can be as deep as 12" for heavy vehicles and deeper for tracked vehicles. Coarse-grained soils provide better support for several reasons: (a) the strength of coarse soils tends to increase more rapidly with depth than do fine-grained soils, ensuring the critical depth is near the surface; (b) they do not readily remold, meaning the traction of vehicles behind the lead are improved as long as they remain in the ruts formed by the first; (c) the trafficability of coarse-grained soils can be improved by reducing tire pressure (distributing the vehicle weight over a wider area) more so than can trafficability in fine-grained soils.

Very fine-grained soils pose an additional problem affecting visibility; they can be readily airborne. Significant loading of the atmosphere by windblown dust occurs for particles smaller than $20\ \mu\text{m}$ (clay-size particles that are common in low-lying areas), which are easily lofted by vehicular and foot traffic. While this fine-grained dust may be found exposed on the surface such as on playas, it can also be present in large quantities in the near subsurface, commonly occurring as several centimeter thick deposits immediately under, and protected from erosion but not from vehicular traffic, in geologic terrains called “desert pavements” commonly found on alluvial fans.

Soil composition and moisture

Fine-grained soils can also contain a significant component of clay minerals, which because of their composition are both hygroscopic (stay wet) and often expansive and tacky when wet, compounding the hazard of their traversing. Furthermore, a wet subsurface is not easy to discern. It can often be camouflaged by a thin dry top surface, especially in dry lake beds and sabkhas. Together, these characteristics multiply the hazard posed by moist soils. Soil moisture content begins to dramatically degrade trafficability at $\sim 20\%$, with a soil becoming liquid at $\sim 40\%$ volumetric water. Clay minerals can absorb many percent of water interstitially (within themselves) and upon compressing when traversed will tend to release this water, increasing the amount of available water in the surface further degrading trafficability.

Relevant desert terrain types

It is important to distinguish specific desert terrain types because the trafficability of a terrain is quite dependent

on its geology, which controls the major factors of soil strength: grain size, water content, and composition. Mountainous deserts often possess several different geologic provinces such as active alluvial fans (including dissected fans, washes, alluvial terraces), rocky highlands, and pediments. Badlands, alluvial plains, playas, and sand dunes are also common in many deserts.

Dry lakes

Playas and variants such as alkali flats, sabkhas, mud flats, and salt flats result from the evaporation of closed bodies of water and pose significant hazards because of both fine grain, sticky clay, and commonly a wet near subsurface. Sabkhas are supratidal salt flats, and thus form near coastlines, and are common in the Persian Gulf area. Like playas, mature sabkhas have standing water only after intense rain storms. All these evaporate landforms, though, can have wet soils (mud, salt) immediately under a dry crust easily broken through by all vehicle types. A wet near surface can persist due to groundwater recharge by continental waters. With a slope of $<1\%$ and slow evaporation due to the hygroscopic nature of the salt and clay-rich soil, groundwater is almost always present within 2 m of the surface, and thus, these surfaces can be more hazardous to tracked vehicles than wheeled vehicles. However, once compacted, the soils do not absorb water as well; thus, unimproved roads can provide a safe traverse when otherwise impassible. These landforms also accumulate significant very fine-grained windblown siliceous and salt-rich dust, easily lofted by traffic or wind.

Alluvial fans

These are apron-like deposits of material eroded from the mountains at their base. They can often be found adjacent to playas. In general, they are of poorly sorted material that is carried down in flash floods from the mountains. Their horizontal extent is on par with their mountainous source regions, and individual fans can merge to form composite landforms extending hundreds of kilometers. Surface roughness ranges from impassible boulder-strewn fields to smooth gravel plains. Fans are also often dissected by internal drainage channels requiring a meandering path to navigate. The horizontal profile of a fan is characteristic of the dominant size of material. Fans of large stones and boulders have a decidedly convex profile and are thus rough and difficult to traverse. These tend to be active young fans. Older, mature fans are broader and flatter, often covered by surfaces of interlocking flat-lying small stones (a geologic feature called “desert pavements”). These “pavements” are easily broken through by all vehicle types exposing a very fine-grained soil, again easily airborne, and if wet, potentially difficult to traverse.

Badlands/Malpais

Badlands are formed through the erosion of sedimentary rocks and soils by wind and water leaving a series of plateaus and steep valleys that can be often navigated.

They should not be confused with Malpais, which are heavily eroded volcanic flows, with very angular stones and a surface roughness at vehicle and pedestrian scales making them difficult or impossible to traverse.

Dunes/dune fields

These unconsolidated sands of quartz, gypsum, and carbonates are deposited through aeolian processes. The largest dunes and fields are longitudinal, oriented parallel to the prevailing wind direction. Each can be hundreds of kilometers long and hundreds of meters high, forming fields of dunes over a hundred kilometer wide. The size and shapes of the grains controls the stability of these surfaces, and trafficability can vary dramatically.

Remotely determining trafficability

Measurement techniques to determine relevant soil properties range from passive to active sensors, from radar to acoustics, and from simple one-parameter measurements to complex radiative transfer modeling. Approaches include thermal imaging, multi- and hyperspectral thermal infrared, polarimetry at optical wavelengths, active and passive radar, acoustics, Vis-SWIR spectroscopy, multi-angle look angles, and [LiDAR](#).

Subsurface moisture

Subsurface moisture can be directly detected by ground-penetrating radar or inferred through observations of the surface. Multiwavelength active interrogation at radar wavelengths offers a proven method for discriminating soil moisture from effects such as volume scattering to develop a vertical profile of moisture to ~1 m depth. These approaches enable the discrimination of moisture from surface roughness. Passive long-wavelength infrared observations can detect relative warming and cooling rates to potentially discriminate areas of near-surface water from dry areas (they warm and cool at different rates). In the future, it may also be possible to employ sound for the interrogation of subsurface moisture. The acoustical response of a soil depends directly upon its strength and thus potentially offers a direct method for measuring the trafficability of a surface.

Soil grain size

Specific soil grain size can be robustly determined for well-sorted soils only because of a strong qualitative relationship between grain size and brightness in the near and shortwave infrared given a uniform composition. Additional effects due to subpixel surface roughness also need to be considered. The presence of clay-sized particles can be inferred more reliably due to compositional signatures, and clay minerals can be identified via specific optical characteristics. However, because clay minerals also exist as coatings on larger grains, the presence of these minerals does not necessarily indicate a significant abundance or a traffic hazard. Landforms known to contain fine-grained materials – desert pavements and dry lakes – can

themselves be detected by compositional signatures and morphology. Dry lake beds contain significant hydrated salts that have unique optical signatures in the infrared. The closely packed stones composing the surface of a desert pavement, under which several centimeters of fine-grained dust often accumulates, also have unique optical and radar reflectance and polarization properties.

Surface roughness

Several techniques exist to quantify the roughness of a surface. Direct imaging with active techniques such as microwave radar and LiDAR can provide absolute measurements, although resource intensive. However, passive microwave methods provide too coarse spatial resolution to provide useful trafficability information. Surface roughness can also be estimated via passive nadir and multi-angle optical measurements at visible and infrared wavelengths, though less precisely, with a large number of measurements potentially enabling a full reflectance function characterization to infer absolute grain size and with relative grain size information possible from the measurement of as few as two angles.

Summary

The trafficability of desert terrain is controlled by soil strength, surface roughness, and the propensity for a surface to generate dust (dust loading) that would obscure vision or impair engine performance. It is the combination of soil strength and vehicle footprint (the area over which the weight of the vehicle is distributed) that dominates a surface's trafficability. Perhaps the most difficult to recognize, and likely the most important factor for trafficability in desert terrains, is soil strength, which itself is dependent upon several factors that can be measured or estimated: soil moisture, grain (or clast) size, and composition. Subsurface moisture is particularly hazardous because dry surfaces in many terrains (e.g., playas, sabkhas, desert pavements) often disguise wet subsurfaces.

Bibliography

- Anderson, K., and Croft, H., 2009. Remote sensing of soil surface properties. *Physical Geography*, **33**, 457, doi:10.1177/0309133309346644.
- Army Field Manual 5-430, 1993. Chapter 7. Soils Trafficability.
- Army Field Manual 90-3/FMFM 7-27, 1990. Desert Operations.
- Caldwell, T. G., McDonald, E. V., Bacon, S. N., and Stullenbarger, G., 2008. The performance and sustainability of vehicle dust courses for military testing. *Journal of Terramechanics*, **45**(6), 213–221.
- Elachi, C., Blom, R., Daily, M., Farry, T., and Saunders, R. S., 1980. Radar imaging of volcanic fields and sand dune fields: implications for VOIR. In *Proceedings of the Radar Geology Workshops*. Snowmass, CO: Jet Propulsion Laboratory, Vol. 80–61, pp. 114–150.
- Hibbitts, C. A., and Gillespie, A., 2008. Polarization of visible light by desert pavements. *Remote Sensing of Environment*, **112**, 1808–1819.
- Hinds, W. C., 1999. *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles*. New York: Wiley.
- Hornback, P., 1988. *The wheeled versus tracked dilemma*. Armor, pp. 33–34.

- King, W. C., Gilewithe, D., Harmon, R. S., McDonald, E., Redmond, K., and Gillies, J., 2004. *Scientific Characterization of Desert Environments for Military Testing, Training, and Operations*. Research Triangle Park, NC: Army Research Office Report to Yuma Proving Ground, Natural Environments Test Office.
- Mushkin, A., and Gillespie, A. R., 2005. Estimating sub-pixel surface roughness using remotely sensed stereoscopic data. *Remote Sensing of Environment*, **99**(1–2), 75–83.
- Ulaby, F. T., Dubois, P. C., and Zyl, J., 1996. Radar mapping of surface soil moisture. *Journal of Hydrology*, **184**, 57–84.

Cross-references

[Geomorphology](#)
[Land Surface Emissivity](#)
[Land Surface Roughness](#)
[Land Surface Temperature](#)
[Soil Moisture](#)
[Soil Properties](#)

TROPOSPHERIC WINDS

Chris Velden
 University of Wisconsin, CIMSS, Madison, WI, USA

Synonyms

Atmospheric motion vectors; Cloud drift winds; Cloud motion winds; Satellite-derived winds; Water vapor motion winds

Definition

Troposphere. The lowest layer of the atmosphere between the Earth's surface and extending up to ~10–15 km, characterized by decreasing temperature with increasing altitude, and where most weather exists.

Tropospheric winds. Tropospheric winds can be estimated from atmospheric motion vectors, which are derived by tracking the displacement of clouds or water vapor in a sequence of images taken from meteorological satellites.

Introduction

Meteorological satellites perform a variety of functions. One of the primary derived products from the imagery they provide is atmospheric motion vectors (AMVs). These vectors are obtained by tracking the displacement of cloud or water vapor features in sequential images in order to deduce an estimate of the horizontal wind at that location. AMV extraction methods have evolved to the point of becoming fully automated and precise, and data sets are routinely produced by several national operational data providers around the globe. AMVs derived from meteorological satellite imagery have become an important source of tropospheric wind information, especially over the oceans where routine observations from other sources are scarce.

AMV extraction methods have been under development since the late 1960s (Menzel, 2001). Initial efforts

required human interaction to target and track cloud features in satellite imagery displayed with the Man computer Interactive Data Access System (McIDAS), developed by the Space Science and Engineering Center (SSEC) at the University of Wisconsin. These methods were very labor intensive and subjective, thus preventing the data from being used operationally. Since the late 1980s, automated techniques have been developed and advanced to the level where operational production was possible, and significant impact on global forecast models has been achieved. Tropospheric winds derived from operational algorithms are now being utilized quantitatively in global numerical models worldwide.

AMV extraction methods have consistently improved through collaborative research efforts which have positively impacted the processing strategies and algorithm functionality. The International Winds Working Group (IWWG) sanctioned by the World Meteorological Organization was established in 1991 and has facilitated the exchange of AMV research and improved algorithms. These upgraded algorithms have been implemented operationally by collaborating international meteorological agencies over the years. This has resulted in global availability and application of “state-of-the-art” satellite-derived AMV algorithms that have impacted many different areas of operational and research meteorology (Velden et al., 2005).

Data processing overview

The basic AMV derivation process involves the execution of a sequence of steps: image navigation/registration, target selection, target height assignment, vector displacement determination, and quality control. Prior to execution of the AMV derivation process, two input data sets must be collected and staged. The first is a sequence of three or more satellite images. The optimal time resolution between images is dependent on the spatial and spectral resolution of the imagery and the domain of the application. In practice, for most applications, this is ~15 min. The second data set required is a set of numerical model forecast fields valid for the image times being processed. These fields help guide the altitude placement of the derived vectors.

Once the input data are properly staged, AMV processing can begin. Potential tracking targets can be well-defined cloud features or sharp gradients in water vapor imagery. Target height assignment values are calculated using a number of different methods depending upon the spectral nature of the imagery being utilized. Targets (features) are then tracked between successive image pairs, looking for high correlations. If found, displacements are calculated to determine AMVs (estimated tropospheric winds at those locations). At this point, each AMV is quality controlled using a series of tests that check for consistency and fit of the vector with its neighboring data. Various quality flag values are assigned from this analysis to each wind vector examined and passed on to the user community.

Applications

Numerical weather prediction

Perhaps the most important application of AMVs is their assimilation as tropospheric winds into numerical forecast models. Proper analysis of atmospheric conditions at the model initialization time is paramount to subsequent accurate forecasts. The AMVs provide critical data over regions normally void of observations from other sources. The positive impact of the AMV data on numerical model forecasts has been proven in many studies, over many meteorological scales (Velden, 1996; Goerss et al., 1998; Soden et al., 2001; Rohn et al., 2001; LeMarshall et al., 2002; Kelly, 2004). Tropospheric winds derived from operational AMV algorithms are now being utilized quantitatively in global numerical models worldwide.

Tropical cyclones

The proper specification of the environmental flow field in the vicinity of tropical cyclones (TCs) can determine the storm motion, forecast track, and intensity. Since TCs are primarily oceanic events, satellite observations are critical to monitor the evolving environmental wind conditions. Fields of AMVs can be utilized to determine TC steering flows (Velden et al., 1992, 1998; Bosart et al., 1999; Dunion et al., 2002) and also to deduce the local vertical wind shear which is crucial to TC intensity modulation.

Polar regions

Geostationary satellites do not generally provide useful AMV information poleward of about 65° latitude due to their equatorial orbit. Polar-orbiting satellites have orbital periods of approximately 100 min and, therefore, provide consecutive scans over nearly the same polar regions at these intervals. Imagers on polar orbiters have scanning swath widths of 2,300 km or more, providing significant geographic overlap by consecutive orbits at high latitudes. These temporal and spatial characteristics provide the opportunity for deriving AMVs from successive orbits (single or multiple satellites) over the polar regions. The AMV processing methodology employed is based on the automated tracking algorithms used with geostationary satellites, modified for use with the Moderate Resolution Imaging Spectroradiometer (MODIS) on board the National Aeronautics and Space Administration's (NASA's) polar-orbiting *Terra* and *Aqua* satellites. Cloud and water vapor tracking with MODIS data is described further in Key et al. (2003). Numerical model impact studies have shown that the MODIS AMV data can have a positive impact on high-latitude forecasts. The AMVs are now routinely assimilated operationally at most global modeling centers.

Mesoscale and rapid scanning

Weather forecasters can utilize the additional detail that can be captured from more frequent imaging in events associated with rapidly changing cloud structures.

Successful tracking of features such as cumulus clouds over land, which have lifetimes that can be considerably less than 15–30 min, demands the use of imagery with time intervals in the 5 min range. Results from a study by Velden et al. (2000) showed that more frequent image sampling improved the AMV field density and quality, and the optimal tracking time interval was dependent on the spectral band employed. Fortunately, the global satellite providers and AMV producers are realizing the gains that are possible from rapid scans. So, the outlook is good for more frequent scanning cycles on near-term future geostationary satellite systems worldwide. As rapid scan imaging becomes a routine part of observing system strategies, quantities such as operational AMVs will greatly benefit. This, in turn, will directly benefit numerical model predictions and the operational forecast community.

Summary and outlook

The scientific community working on the retrieval of tropospheric winds from meteorological satellites recognizes the importance of steadily improving AMV observational quality so that the data continue to make a positive contribution to advancing data assimilation and forecast systems. In order to keep pace with these demands, innovative research toward improving ways of deriving winds from satellites has been a focus of the World Meteorological Organization and Coordination Group for Meteorological Satellites cosponsored International Winds Workshops. These workshops are held every 2 years and bring together AMV researchers from around the world to present new, innovative ideas on AMV extraction techniques, interpretation, and applications (Velden et al., 2005). The NWP community is always well represented at these workshops and provides an important exchange of information on the latest in data assimilation issues. The AMV applications briefly described here represent just a small sample of the new innovations in satellite-produced wind technologies, derivation methodologies, and products that have recently become available. The AMV data processing community is continuing to progress in all of these areas and is looking forward to pending new opportunities with emerging advanced sensor technologies.

Bibliography

- Bosart, L. F., Velden, C. S., Bracken, W. E., Molinari, J., and Black, P. G., 1999. Environmental influences on the rapid intensification of hurricane Opal (1995) over the Gulf of Mexico. *Monthly Weather Review*, **128**, 322–352.
- Dunion, J., Houston, S. H., Velden, C. S., and Powell, M. D., 2002. Application of surface adjusted GOES low-level cloud-drift winds in the environment of Atlantic tropical cyclones. Part II: integration into surface wind analyses. *Monthly Weather Review*, **130**, 1347–1355.
- Goerss, J. S., Velden, C. S., and Hawkins, J. D., 1998. The impact of multispectral GOES-8 wind information on Atlantic tropical cyclone track forecasts in 1995. Part II: NOGAPS forecasts. *Monthly Weather Review*, **126**, 1219–1227.

- Kelly G., 2004. Observing system experiments of all main data types in the ECMWF operational system. In *3rd WMO Numerical Weather Prediction OSE Workshop*. Alpbach, Austria: WMO. Technical Report 1228, pp. 32–36.
- Key, J., Santek, D., Velden, C. S., Bormann, N., Thepaut, J.-N., Riishojgaard, L. P., Zhu, Y., and Menzel, W. P., 2003. Cloud-drift and water vapor winds in the polar regions from MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 482–492.
- LeMarshall, J., Mills, G., Seecamp, R., Pescod, N., Leslie, L., and A. Rea, 2002. High density AMVs and their application to operational NWP. In *Proceedings of the Sixth International Winds Workshop*. Madison, WI: WMO, pp. 137–146.
- Menzel, W. P., 2001. Cloud tracking with satellite imagery: from the pioneering work of Ted Fujita to the present. *Bulletin of the American Meteorological Society*, **82**, 33–47.
- Rohn, M., Kelly, G., and Saunders, R. W., 2001. Impact of a new cloud motion wind product from METEOSAT on NWP analyses and forecasts. *Monthly Weather Review*, **129**, 2392–2403.
- Soden, B. J., Velden, C. S., and Tuleya, R. E., 2001. The impact of satellite winds on experimental GFDL hurricane model forecasts. *Monthly Weather Review*, **129**, 835–852.
- Velden, C. S., 1996. Winds Derived from Geostationary Satellite Moisture Channel Observations: Applications and Impact on Numerical Weather Prediction. *Meteorology and Atmospheric Physics*, **60**, 37–46.
- Velden, C., Hayden, C. M., Menzel, W. P., Franklin, J. L., and Lynch, J., 1992. The impact of satellite – derived winds on numerical hurricane track forecasting. *Weather and Forecasting*, **7**, 107–118.
- Velden, C., Olander, T. L., and Wanzong, S., 1998. The impact of multispectral GOES-8 wind information on Atlantic tropical cyclone track forecasts in 1995. Part I: dataset methodology, description, and case analysis. *Monthly Weather Review*, **126**, 1202–1218.
- Velden, C., Stettner, D., and Daniels, J., 2000. Wind vector fields derived from GOES rapid scan imagery. In Preprints, *10th Conference on Satellite Meteorology*. Long Beach, CA: American Meteorological Society, pp. 20–23.
- Velden, C., et al., 2005. Recent innovations in deriving tropospheric winds from meteorological satellites. *Bulletin of the American Meteorological Society*, **86**, 205–223.

Cross-references

[Data Assimilation](#)
[Data Processing, SAR Sensors](#)
[Geophysical Retrieval, Overview](#)
[Observational Platforms, Aircraft, and UAVs](#)
[Observational Systems, Satellite](#)
[Weather Prediction](#)

U

ULTRAVIOLET REMOTE SENSING

Arlin Krueger
Atmospheric Chemistry and Dynamics Laboratory
(Code 614), NASA/Goddard Space Flight Center,
Greenbelt, MD, USA

Synonyms

UV

Definitions

Ultraviolet spectral region: Wavelengths less than 400 nm, the short wavelength limit of human vision, and longer than 10 nm, where the x-ray spectrum begins.

UV remote sensing goals: The Earth's ozone layer, volcanic eruptions, and airglow and auroral emissions.

UV spectral divisions: Regions based on properties of atmospheric gases and optical materials:

- Near UV (NUV) 400–300 nm
- Middle UV (MUV) 300–200 nm
- Far UV (FUV) 200–100 nm
- Extreme UV (EUV) 100–10 nm

Introduction

Near and middle UV wavelengths have information about ozone, sulfur dioxide, and trace gases in the troposphere and stratosphere (0–50 km) of interest to the atmospheric and volcanic sciences. Far and extreme UV wavelengths have information about airglow and auroral emissions from the thermosphere (>100 km) of interest in aeronomy and space weather. These studies have developed independently because the first deals with reflected sunlight, while the second deals with emissions from atoms or molecules in excited states. Excellent books on UV remote sensing science and technology are Green (1966) and Huffman (1992).

Near and middle UV remote sensing

The greatest public impact of satellite UV remote sensing of the atmosphere comes from measurements of *ozone*, the constituent that blocks UV sunlight from reaching the biosphere. Early rocket and satellite experiments used UV light to measure ozone. Rocket techniques measure the extinction of MUV sunlight versus height to derive the ozone distribution. Satellite techniques measure backscattered NUV and MUV sunlight to infer the vertical ozone distribution and total amounts. Remote sensing of an atmospheric constituent requires forward model calculations of the scattering and absorption processes for the given observing conditions, coupled with inverse calculations to retrieve the distribution of the constituent from the observed spectrum of light. Three physical processes control the transfer of near and middle UV radiation in the atmosphere: absorption by gases, scattering by molecules and aerosols, and reflection from clouds or the Earth's surface.

Atmospheric UV absorbers

The dominant atmospheric optical process in the NUV and MUV is *absorption* by oxygen and ozone. The absorption bands of ozone begin weakly near 360 nm and then increase exponentially from about 340 nm to a broad peak near 250 nm. Ozone is an unusual atmospheric constituent with an inverted distribution with altitude due to its formation from photodissociation of oxygen at MUV wavelengths. Ozone production is balanced by destruction by radicals in the H, N, O, and halogen families. The ozone density is low in the troposphere, increases in the stratosphere to a maximum around 25 hPa (~25 km height), and then decreases exponentially with height up to 60 km. This unusual distribution influences methods for remote sensing of ozone from space. NUV sunlight reaches the biosphere where two components, UVA and UVB, are recognized for their biologic action. Another

MUV absorber is sulfur dioxide, which has a short lifetime in the troposphere and is normally found only near sources, such as volcanoes, smelters, and fossil fuel power plants. However, large amounts are found in volcanic eruption plumes in the stratosphere where the SO₂ lifetime is about a month. Volcanic plumes can travel global distances before conversion to sulfate molecules.

Atmospheric scattering

The second major optical process at ultraviolet wavelengths is Rayleigh *scattering* by air molecules. This strong, inverse fourth-power wavelength-dependent scattering produces the blue sky. From space, Rayleigh backscattered UV sunlight is utilized to derive the vertical distribution of ozone and the total ozone amount. In addition to Rayleigh scattering, inelastic rotational Raman scattering shifts the wavelength of the light. This has an important effect on scattered sunlight due to Fraunhofer structure in the solar spectrum. Thus, spectral dips in Earth radiance are filled and peaks are diminished.

Reflection from the surface and clouds

NUV light that reaches the Earth's surface is reflected from the ground, vegetation, and water. The reflectivity of all terrestrial surfaces is small (<10 %), but clouds are very reflective (80–100 % for thick clouds) and can make a large contribution at the top of the atmosphere. Total column ozone amounts are derived using a radiative transfer model to account for the absorption, scattering, and reflections.

Radiative transfer models

Calculation of the radiance of the atmosphere is a major component of UV remote sensing. Numerical models of the radiance of the Earth's atmosphere were developed by Dave (1965) who computed tables of radiance versus geometry, wavelength, ozone profile, and surface reflectivity for use in satellite retrieval of total ozone. Other RT codes, such as LOTRAN and HITRAN, are available for general use.

UV remote sensing from the surface, balloons, and rockets

Total column ozone amounts have been measured since the 1920s through differential absorption of sunlight at NUV wavelengths using a double quartz prism spectrophotometer developed by G.M.B. Dobson (1968) at Oxford University. Dobson spectrophotometers are still in use for monitoring ozone trends at stations across the world. Later, Paetzold and colleagues in Germany developed balloon-borne UV filter photometers to measure the vertical distribution of ozone. Ozone densities are computed from the change in NUV solar irradiance with altitude. Ozone measurements in the photochemical production region above 30 km required rockets. Johnson et al. (1952) launched photographic spectrographs on Aerobee rockets for the first definitive data on

photochemical ozone densities up to 70 km. Krueger (1973) later developed MUV filter photometers for meteorological sounding rockets to measure latitudinal and seasonal changes and for calibration of satellite ozone profile data. The US Standard Atmosphere 1976 midlatitude ozone distribution is based on many of the early rocket soundings (Krueger and Minzner, 1976).

UV remote sensing from satellites

Early space experiments used UV filter photometers for measurements of the backscattered radiance of the sunlit Earth (e.g., Friedman et al., 1963). The first spectrum of backscattered MUV sunlight was measured with a spectrometer (e.g., Iozenas et al., 1969). These early experiments produced data to evaluate *forward* calculations from radiative transfer models. However, to remotely sense the ozone distribution that produced the observed radiance requires an *inversion* algorithm and a radiative transfer model to derive the *path* of the photons into and out of the atmosphere.

Vertical ozone distribution

The first ozone profile experiment measured the extinction of 260 nm sunlight with a filter photometer as rays from the Sun passed through the limb of the Earth to the satellite (Rawcliffe et al., 1963). The photon path in this case is simply the geometry of rays passing through the limb as a function of minimum ray altitude accounting for refraction and the finite size of the solar disc. This method is limited to one sunrise and sunset ozone profile during each orbit, but, similar to rocket soundings, can accurately be calibrated from rays taken above the ozone layer.

A more general method for remote sensing of ozone profiles from space involves measuring scattered sunlight from the nadir at strongly absorbed wavelengths between 250 and 300 nm where sunlight is totally absorbed by **stratospheric ozone** if it is not scattered back to space. The backscattered light originates only from altitudes above the depth for total absorption at each wavelength. As the ozone absorption coefficient increases, the penetration depth decreases. Thus, a nadir viewing instrument can determine the ozone profile by measuring the albedo (ratio of Earth radiance to solar irradiance) at several wavelengths. The path traveled by photons is no longer geometric, so a radiative transfer model is required to derive the ozone distribution. At upper stratospheric altitudes (25–50 km), the air densities are low enough that a solar photon is likely to be scattered only once in its path back to space. Under these conditions, an analytic algorithm was devised (Twomey, 1961) to retrieve the upper level ozone distribution from albedos at a few MUV wavelengths. The first upper level ozone profiles were obtained using MUV spectra from the OGO-4 satellite in 1967 (Anderson et al., 1969) using the Twomey algorithm and a coincident rocket ozone sounding for calibration. At altitudes below the ozone maximum near 25 km, multiple

scattering becomes important and only statistically constrained profile solutions are possible.

Global ozone monitoring

Fundamental problems for satellite spectrometers are the high dynamic range of signals across the UV spectrum, stray light control, and a solar calibration. The first satellite instrument built specifically to deal with these problems was the NASA Backscatter Ultraviolet (BUV) double monochromator proposed by J. V. Dave and C. L. Mateer of the National Center for Atmospheric Research and developed by D. F. Heath of the Goddard Space Flight Center (GSFC) for flight on the Nimbus 4 satellite in 1970. Twelve optimally selected wavelengths were used for MUV ozone profile retrieval and to test a new NUV total ozone algorithm (Dave and Mateer, 1967). This flight produced global upper level ozone data (Heath et al., 1973) and demonstrated that BUV total ozone retrievals were accurate (Mateer et al., 1971) using a multiple-scattering radiative transfer model (Dave, 1965) to determine the path for multiply scattered NUV photons.

Ozone profiles

BUV ozone profiles provided the first daily global information in the photochemical regime as a function of latitude and season. In the 4 year BUV record, the most significant scientific finding was that stratospheric ozone was kept in steady state by the catalytic cycles found in laboratory experiments. In August 1972, ozone losses appeared in the BUV data in the auroral zone when a solar proton event produced large quantities of nitric oxide in the only locations where the Earth's magnetic field permitted protons to reach the stratosphere (Heath et al., 1977). The success led to an advanced Solar Backscatter Ultraviolet (SBUV) instrument on the Nimbus 7 satellite in 1978 (Heath et al., 1975) and later to operational use on NOAA satellites. GSFC scientists undertook a major algorithm development effort to extend the profile retrieval below the single-scattering domain by accounting for multiple scattering and constraining the solution using maximum likelihood theory (Bhartia et al., 1996).

Total ozone

A new Total Ozone Mapping Spectrometer (TOMS) instrument was proposed by A. J. Krueger and flown with the SBUV on Nimbus 7 to ascertain the cause of high variability (Krueger, 1989). The instrument field of view (50×50 km at nadir) was scanned across track as the satellite moved in its orbit to generate full daily global maps of ozone. Total ozone retrievals use NUV wavelengths where sunlight penetrates to the Earth's surface, so that the entire column can be measured. The solution depends on accurate radiative transfer calculations for a simulated atmosphere containing climatological ozone profiles, observing geometry and reflections from the surface and

clouds (Dave, 1965). The instrument and algorithms are documented in <http://ozoneaq.gsfc.nasa.gov/n7sat.md>.

The TOMS total ozone images showed wave structure that matched planetary waves, demonstrating the control of atmospheric dynamics on total ozone through vertical displacement of air in the lower stratosphere. Figure 1 is an example of planetary waves in the total ozone field on April 16, 1984. This dynamic control of total ozone structure prevailed until chemistry was shown to produce major ozone features at a time when environmental concerns were expressed about chlorine in the atmosphere. A hole in the ozone over the Antarctic was found each October in TOMS data (e.g., Stolarski et al., 1986). This polar ozone depression began deepening in the 1980s in response to growth of chlorine from chlorofluorocarbons (CFC) released in the atmosphere. Chlorine is an ozone catalyst that is normally sequestered in stable compounds. However, surface reactions on aerosols in the wintertime Antarctic lower stratosphere produce chlorine molecules that are released as chlorine atoms when the Sun rises in spring, decreasing total ozone to one third of pre-CFC levels. Dramatic images of the Antarctic Ozone Hole (Figure 2) contributed to an international agreement to phase out production of CFC compounds: the Montreal Protocol on Substances that Deplete the Ozone Layer. Systematic TOMS global coverage is ideal for monitoring total ozone trends to validate models of global ozone change. Stabilizing the calibration of TOMS to detect trends of 1 %/decade is challenging (McPeters et al., 1996).

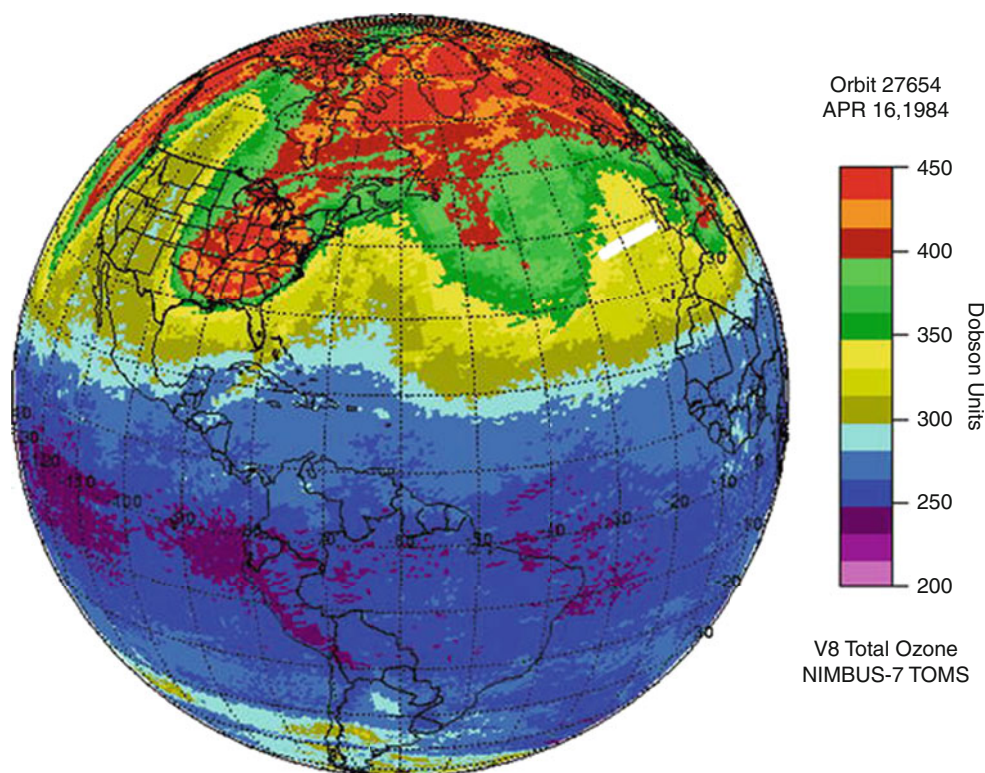
Sulfur dioxide and volcanic eruptions

A geophysical bonus is volcanic sulfur dioxide found in TOMS data when very large amounts appeared in the 1982 eruption of the Mexican El Chichon volcano. Discrimination between ozone and sulfur dioxide required a new algorithm (Krueger et al., 1995). The TOMS spatial mapping design proved to be excellent for measurement of transient phenomena. The size of all volcanic eruptions in the 25 year TOMS data record was determined (Carn et al., 2003) for climate research.

Absorbing aerosols

A third geophysical product from the TOMS UV radiances is an aerosol index (AI), a measure of dust, smoke, volcanic ash, and other absorbing aerosols (Torres et al., 1998). Aerosols change the spectral dependence of backscattered light at nonabsorbed wavelengths. TOMS AI is of particular value over land because of the low UV reflectivity of all terrestrial surfaces. Saharan dust is tracked from its desert origin to the Caribbean islands, while smoke can be tracked from forest fires in Canada to the Atlantic Ocean.

Early satellite ozone experiments are summarized in Krueger et al. (1980). Reviews of TOMS ozone and sulfur dioxide findings are in Krueger (1995) and Krueger et al. (2000). Three additional TOMS missions on research satellites and a series of operational SBUV flights on NOAA satellites continued the ozone and volcano eruption record for over 25 years.



Ultraviolet Remote Sensing, Figure 1 The horizontal distribution of total ozone on April 16, 1984, from TOMS data. Total ozone is measured in Dobson Units (DU) or milli-atm.-cm, the thickness of a layer of pure ozone at NTP conditions equal to all the overhead ozone. Tropical total ozone is always low, near 250 DU (2.5 atm.-mm), but higher latitude ozone varies seasonally, reaching 500 DU (5 atm.-mm) in springtime. The large waves are produced by upper troposphere meteorological troughs which bring stratospheric air down and create total ozone and ridges which lift and destroy total ozone. The jet streams follow the ozone contours.

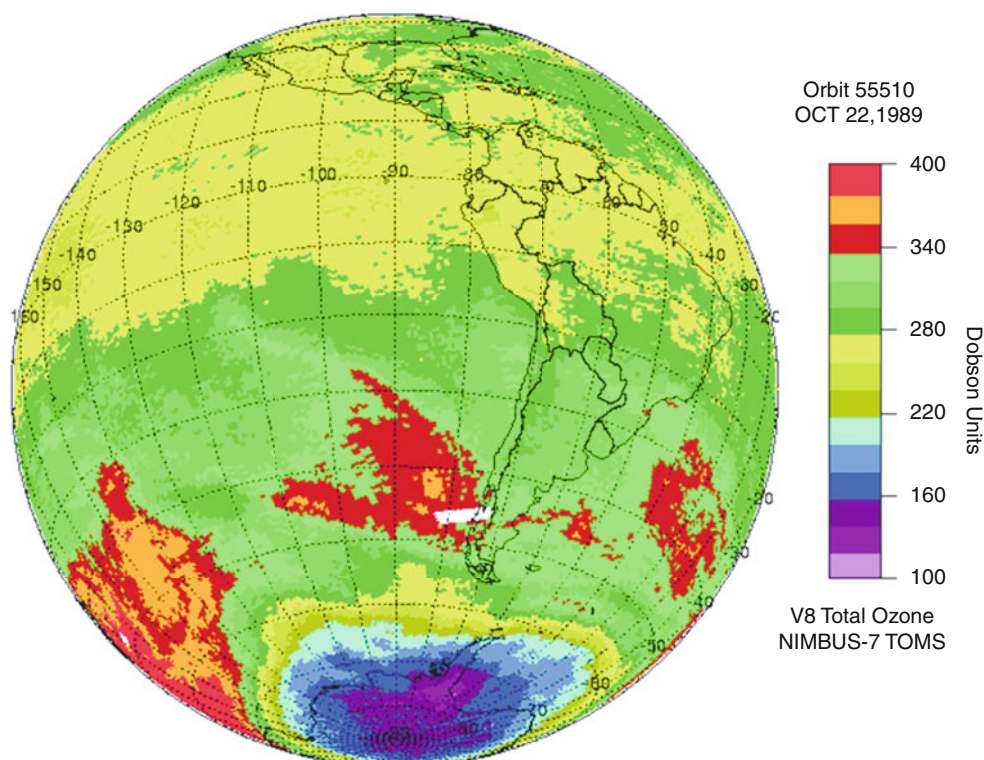
Second-generation satellite UV remote sensing instruments

In 1990, with the concern about ozone depletion, the European Space Agency accepted a proposal from J. Burrows of the Max Planck Institute/Mainz for a new UV/visible remote sensing instrument for flight on the ERS-2 spacecraft. By this time, spacecraft telemetry had advanced so that far more information could be collected. The Global Ozone Monitoring Instrument (GOME) was developed to measure the full spectrum from 240 to 790 nm (Burrows et al., 1999). This allows absorbing species to be positively identified and wavelength errors and rotational Raman scattering to be corrected. The extended spectral coverage allows detection of other trace gases such as NO_2 , BrO, OClO, and formaldehyde (HCHO) in addition to ozone, sulfur dioxide, and aerosols. The first measurements of these trace gases from space were made with GOME data.

GOME algorithms use the Differential Optical Absorption Spectroscopy (DOAS) method that is less dependent on absolute radiance calibrations than are the BUV algorithms. DOAS uses a high-pass filter to isolate spectral band features in the radiance data due

to the absorption spectrum of the absorber, using a spectral interval broad enough to encompass band structure. The slant-path absorber amount is derived by a least squares minimization of the difference between observations and lab spectral data. An effective optical path for the selected band is computed using a radiative transfer model, similar to the BUV model. This path depends on wavelength, absorption coefficients, geometry, reflectivity of the ground or clouds, and the vertical distribution of all absorbers in the selected wavelength band.

GOME was followed in 2002 by the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) research instrument launched on ENVISAT. SCIAMACHY added limb scattered light observations for profiles of stratospheric constituents. Nadir spatial resolution was improved over GOME with a ground resolution of 60×30 km and a swath width of 960 km. An online book, SCIAMACHY: Monitoring the Changing Earth's Atmosphere, is available: http://atmos.caf.dlr.de/projects/scops/sciamachy_book/sciamachy_book.html. Many of the scientific products are available through <http://www.iup.uni-bremen.de/sciamachy/index.html>.



Ultraviolet Remote Sensing, Figure 2 The Antarctic Ozone Hole on October 22, 1989, viewed in TOMS total ozone data. The deep oval region over Antarctica is produced by destruction of ozone in the lower stratosphere by chlorine in the polar vortex beginning in September and lasting until breakup of the vortex in November. Within the vortex, total ozone has decreased significantly since 1980 due to the growth of chlorofluorocarbons. Unprecedented low amounts less than 100 DU have been found. The midlatitude air surrounding the vortex remains unaffected with values above 400 DU because it is isolated from the polar air.

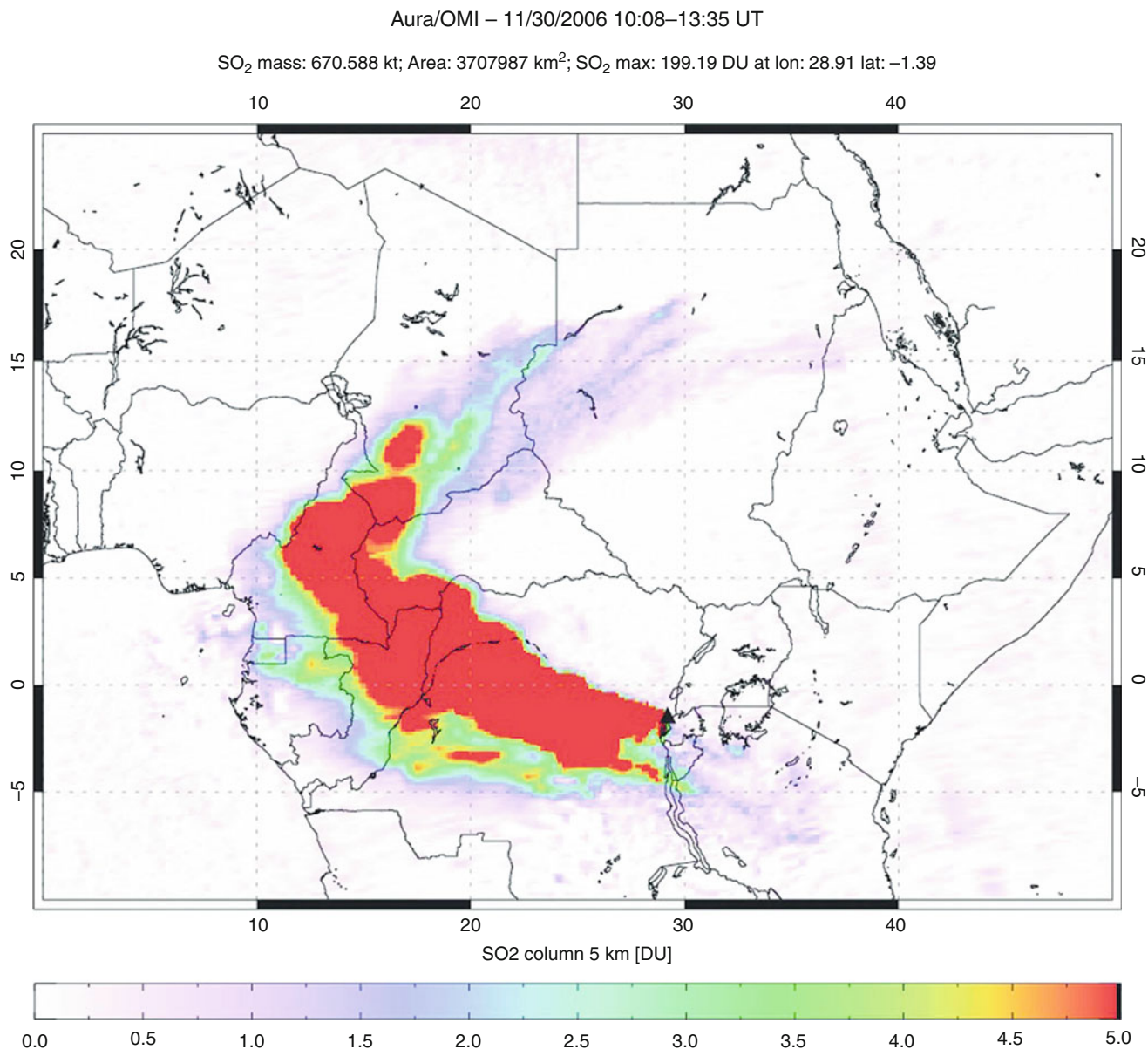
The Dutch-Finnish Ozone Measuring Instrument (OMI) (Levelt et al., 2006) launched in 2004 on the NASA EOS Aura platform (<http://aura.gsfc.nasa.gov/instruments/omi/index.html>; <http://www.knmi.nl/omi/research/science/>) returned to the TOMS concept of contiguous daily global coverage but with an eightfold increase in ground resolution (13×24 km at nadir) and with full spectral coverage from 270 to 500 nm. The contiguous coverage resulted in exceptional daily images of NO_2 and SO_2 sources. Figure 3 illustrates the SO_2 from an eruption of Nyamuragira volcano on November 30, 2006 (<http://so2.gsfc.nasa.gov/>). A 20-fold improvement in SO_2 sensitivity over TOMS allows monitoring of volcanic degassing and pollution from smelters and fossil fuel burning.

A second-generation GOME instrument, GOME-2, was launched on the first European polar-orbiting operational meteorological satellite, MetOp-A, in 2006. This was followed in 2012 with a second GOME-2 instrument on the MetOp-B satellite. GOME-2 adds a spatial scanner with ground resolution similar to TOMS. The series of MetOp flights will provide a continuation of the long-term record of ozone trends. The GOME-2 instrument and missions are described in http://www.esa.int/esaLP/SEMTEG23IE_LPmetop_0.html.

The USA launched its first hyperspectral UV instrument, Ozone Mapping and Profiler Suite (OMPS) (<http://npp.gsfc.nasa.gov/omps.html>), on the Suomi National Polar-orbiting Partnership (NPP) satellite in 2011 (<http://npp.gsfc.nasa.gov/index.html>). The NPP OMPS is the first of a series planned for Joint Polar Satellite System (JPSS) operational missions (<http://www.jpss.noaa.gov/>). OMPS continues ozone trend monitoring that began with the Nimbus 7 SBUV/TOMS mission in 1978 with the same observing requirements except that limb observations are added for improved vertical resolution of ozone profiles.

Far and extreme UV remote sensing

Remote sensing at FUV and EUV wavelengths views emissions from the upper atmosphere that result from radiative decay of molecules and atoms in excited states. Chemical reactions and solar photons produce airglow, while energetic particles produce the aurora. Models are then used to infer air densities in the thermosphere and the energy input from solar photons and energetic particles. UV observations are part of Global Space Weather Systems, in development to predict communication



Ultraviolet Remote Sensing, Figure 3 The plume of sulfur dioxide from an eruption of Niyamuragira volcano on November 30, 2006, from OMI data. Sulfur dioxide is distinguished from ozone because the absorption spectra differ in detail. The mass of SO₂ in the plume is determined by integrating across the cloud (e.g., 0.67 Tg, in this plume). This is an example of hundreds of volcanic eruptions that have been analyzed from TOMS and OMI data since 1978.

conditions and radiation environments for space travel. Many experiments have been flown on rockets and satellites since 1965 to measure the spectra and spatial distribution of the emissions. The aeronomy experiments and physical processes are discussed in detail in Huffman (1992).

Airglow

At wavelengths less than 200 nm, few solar photons are backscattered to space due to absorption by oxygen.

However, other processes produce emissions, known as *airglow*, both in day and at night, at the low air densities of the upper atmosphere. Molecules may *fluoresce* when stimulated by solar photons and atoms may undergo resonance reradiation. Absorption of UV sunlight can lead to photodissociation of molecules. Photoionization from UV sunlight produces energetic electrons that collide with molecules. These processes and reactions produce excited molecular and atomic states that radiate light at discrete wavelengths. The lowest altitude for airglow emissions is set by the air density where energy is

removed by collisional deexcitation rather than by radiation. Dayglow emissions produce information about N_2 , N, and O densities. Nightglow emissions are primarily from O_2 and NO. Visible nightglow emissions have been monitored from the ground for many years, but ultraviolet lines can only be observed from space. Excellent reviews of airglow and remote sensing are provided by Barth (1966), Meier (1991), and Huffman (1992).

Aurora

Aurorae are spectacular displays of curtains and filaments of green or red light visible from polar latitudes during periods of enhanced solar activity (see, e.g., Akasofu, 1989). They are produced when energetic electrons and protons impact the atmosphere to generate excited states in oxygen and nitrogen atoms and molecules that decay with characteristic spectral lines. Charged particles can only reach the atmosphere in polar regions where the Earth's magnetic field lines become vertical in the auroral ovals. The rich UV spectra viewed from space contain information about solar activity and air densities.

Global space weather systems

In the twenty-first century, FUV images of the Earth are becoming more important as space travel develops and we become more dependent for communications and navigation on space systems. Programs designed to understand the behavior of the thermosphere/ionosphere are maturing. The NASA POLAR mission (<http://pwg.gsfc.nasa.gov/polar/>), launched in 1996, includes an Ultraviolet Imager to detect auroral emissions in the 130–190 nm region as part of the International Solar Terrestrial Physics Program. The TIMED mission, launched in 2001, includes a Global Ultraviolet Imager (115–180 nm) in a group of instruments to investigate the mesosphere and lower thermosphere (<http://www.timed.jhuapl.edu/WWW/index.php>).

Outlook

UV remote sensing provides robust measurements of atmospheric ozone and sulfur dioxide and is an important component of a global space weather monitoring system. Ozone is produced and held in a steady state in the stratosphere by UV sunlight. The strong absorption limits the biologically damaging, shorter UV solar wavelengths from reaching the surface. The same wavelengths provide the means for measuring the ozone content and vertical distribution from the surface, balloons, rockets, and satellites. Data from the Nimbus 4 Backscatter Ultraviolet (BUV) instrument show conclusively that catalytic processes regulate ozone in the stratosphere. Images from the Total Ozone Mapping Spectrometer (TOMS) instrument on the Nimbus 7 satellite initially provided evidence for meteorological control of total ozone through vertical motions in the lower stratosphere. Later, TOMS images of the Antarctic Ozone Hole provided compelling evidence for large-scale destruction of ozone by chlorine

from industrially produced chlorofluorocarbons and led to international controls of ozone destroying pollutants. Monitoring of ozone trends from UV instruments will be continued for at least 50 more years to verify that destruction by halogens has been mitigated.

Sulfur dioxide is an important climate gas that is produced by volcanoes, smelters, and fossil fuel combustion. Quantitative data on all eruption SO_2 masses is available from the 25 year TOMS dataset. Improved sensitivity from GOME, SCIAMACHY, OMI, and succeeding hyperspectral instruments continues this eruption time series and adds noneruptive degassing from volcanoes, as well as emissions from smelters and in air pollution.

Far UV monitoring of airglow and auroral activity has provided comprehensive information about solar storm effects on the ionosphere and the near-Earth radiation environment. Interruption of communications, navigation, and power transmission has increasingly greater economic consequences in the twenty-first century. As human space travel to the Moon and beyond becomes a reality, this monitoring will be a necessary component of the Global Space Weather System.

Bibliography

- Akasofu, S. I., 1989. The dynamic aurora. *Scientific American*, **199**, 90–97.
- Anderson, G. P., Barth, C. A., Cayla, F., and London, J., 1969. Satellite observations of the vertical ozone distribution in the upper stratosphere. *Annals of Geophysics*, **25**, 341–345.
- Barth, C. A., 1966. The ultraviolet spectroscopy of planets. In Green, A. E. S. (ed.), *The Middle Ultraviolet: Its Science and Technology*. New York: Wiley, pp. 177–218.
- Bhartia, P. K., McPeters, R. D., Mateer, C. L., Flynn, L. E., and Wellemeyer, C., 1996. Algorithm for the estimation of vertical ozone profile from the backscattered ultraviolet (BUV) technique. *Journal of Geophysical Research*, **101**, 18793–18806.
- Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V. V., Ladstädter-Weissenmayer, A., Richter, A., de Beek, R., Hoogen, R., Bramstedt, K.-U., Eichmann, K., Eisinger, M., and Perner, D., 1999. The global ozone monitoring experiment (GOME): mission concept and first scientific results. *Journal of the Atmospheric Sciences*, **56**, 151–175.
- Carn, S. A., Krueger, A. J., Bluth, G. J. S., Schaefer, S. J., Krotkov, N. A., Watson, I. M., and Datta, S., 2003. Volcanic eruption detection by the total ozone mapping spectrometer (TOMS) instruments: a 22-year record of sulfur dioxide and ash emissions. In Oppenheimer, C., Pyle, D. M., and Barclay, J. (eds.), *Volcanic Degassing*. London: Geological Society, Special Publications, Vol. 213, pp. 177–202.
- Dave, J. V., 1965. Multiple scattering in a non-homogeneous, Rayleigh atmosphere. *Journal of the Atmospheric Sciences*, **22**, 273–279.
- Dave, J. V., and Mateer, C. L., 1967. A preliminary study on the possibility of estimating total atmospheric ozone from satellite measurements. *Journal of the Atmospheric Sciences*, **24**, 414–427.
- Dobson, G. M. B., 1968. Forty years' research on atmospheric ozone at Oxford: a history. *Applied Optics*, **7**, 387–405.
- Friedman, R. M., Rawcliffe, R. D., and Meloy, G. E., 1963. Radiance of the upper atmosphere in the middle ultraviolet. *Journal of Geophysical Research*, **68**, 6419–6423.
- Green, A. E. S., 1966. *The Middle Ultraviolet: Its Science and Technology*. New York: Wiley.

- Heath, D. F., Mateer, C. L., and Krueger, A. J., 1973. The Nimbus 4 backscattered ultraviolet (BUV) atmospheric ozone experiment – two years operation. *Pure and Applied Geophysics*, **106–108**, 1238–1253.
- Heath, D. F., Krueger, A. J., Roeder, H. A., and Henderson, B. D., 1975. The solar backscatter ultraviolet and total ozone mapping spectrometer (SBUV/TOMS) for Nimbus G. *Optical Engineering*, **14**, 323–331.
- Heath, D. F., Krueger, A. J., and Crutzen, P. J., 1977. Solar proton event influence on stratospheric ozone. *Science*, **197**, 886–889.
- Huffman, R. E., 1992. *Atmospheric Ultraviolet Remote Sensing*. San Diego: Academic Press Inc.
- Iozenas, V. A., Krasnopol'skiy, V. A., Kuznetsov, A. P., and Lebedinsky, A. I., 1969. Studies of the earth's ozonosphere from satellites. *Atmospheric and Oceanic Physics*, **5**, 149–159.
- Johnson, F. S., Purcell, J. D., Tousey, R., and Watanabe, K., 1952. Direct measurements of the vertical distribution of atmospheric ozone to 70 kilometers altitude. *Journal of Geophysical Research*, **57**, 157–176.
- Krueger, A. J., 1973. The mean ozone distribution from several series of rocket soundings to 52 km at latitudes from 58°S to 64°N. *Pure and Applied Geophysics*, **106–108(1)**, 1272–1280.
- Krueger, A. J., 1989. The global distribution of total ozone: TOMS satellite measurements. *Planetary and Space Science*, **37(12)**, 1555–1565.
- Krueger, A. J., 1995. UV remote sensing of the earth's atmosphere. In *Diagnostic Tools in Atmospheric Physics*. Amsterdam: Ios Press, pp. 155–181.
- Krueger, A. J., and Minzner, R. A., 1976. A mid-latitude ozone model for the 1976 U.S. standard atmosphere. *Journal of Geophysical Research*, **81**, 4477–4481.
- Krueger, A. J., Guenther, B., Fleig, A. J., Heath, D. F., Hilsenrath, E., McPeters, R., and Prabhakara, C., 1980. Satellite ozone measurements. *Philosophical Transactions of the Royal Society of London*, **A296**, 191–204.
- Krueger, A. J., Walter, L. S., Bhartia, P. K., Schnetzler, C. C., Krotkov, N. A., Sprod, I., and Bluth, G. J. S., 1995. Volcanic sulfur dioxide measurements from the total ozone mapping spectrometer instruments. *Journal of Geophysical Research*, **100**, 14057–14076.
- Krueger, A. J., Schaefer, S. J., Krotkov, N., Bluth, G., and Barker, S., 2000. Ultraviolet remote sensing of volcanic emissions. In Mougini-Mark, P. (ed.), *Remote Sensing of Active Volcanism*. Washington, DC: American Geophysical Union. Geophysical Monograph, Vol. 116, pp. 25–43.
- Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J., and Saari, H., 2006. The ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, **44(5)**, 1093–1101. doi:10.1109/TGRS.2006.872333.
- Mateer, C. L., Heath, D. F., and Krueger, A. J., 1971. Estimation of total ozone from satellite measurements of backscattered ultraviolet earth radiance. *Journal of the Atmospheric Sciences*, **28**, 1307–1311.
- McPeters, R. D., Hollandsworth, S. M., Flynn, L. E., and Herman, J. R., 1996. Long-term trends derived from the 16-year combined Nimbus 7/Meteor 3 TOMS version 7 record. *Geophysical Research Letters*, **23**, 3699–3702.
- Meier, R. R., 1991. Ultraviolet spectroscopy and remote sensing of the upper atmosphere. *Space Science Reviews*, **58**, 1–185.
- Rawcliffe, R. D., Meloy, G. E., Friedman, R. M., and Rogers, E. H., 1963. Measurement of vertical distribution of ozone from a polar orbiting satellite. *Journal of Geophysical Research*, **68**, 6425–6429.
- Stolarski, R. S., Krueger, A. J., Schoeberl, M. R., McPeters, R. D., Newman, P. A., and Alpert, J. C., 1986. Nimbus 7 SBUV/TOMS measurements of the springtime Antarctic ozone decrease. *Nature*, **322**, 808–811.
- Torres, O., Bhartia, P. K., Herman, J. R., Ahmad, Z., and Gleason, J. F., 1998. Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation: theoretical. *Journal of Geophysical Research*, **103**, 17099–17110.
- Twomey, S., 1961. The deduction of the vertical distribution of ozone by ultraviolet spectral measurements from a satellite. *Journal of Geophysical Research*, **66**, 2153–2162.

Cross-references

[Stratospheric Ozone](#)

[Trace Gases, Stratosphere, and Mesosphere](#)

ULTRAVIOLET SENSORS

Arlin Krueger

Atmospheric Chemistry and Dynamics Laboratory
(Code 614), NASA/Goddard Space Flight Center,
Greenbelt, MD, USA

Synonyms

Photometers; Spectrometers; UV

Definition

Ultraviolet spectral region. Wavelengths less than 400 nm, the short wavelength limit of human vision, and longer than 10 nm, the limit of X-rays.

UV instrumentation. Differ from “optical” instruments by UV transmitting materials.

UV spectral albedo. The ratio of Earth radiance to solar irradiance as a function of wavelength, comparable to reflectivity.

Introduction

Two classes of sensors are used for ground, balloon, rocket, and satellite UV remote sensing of the Earth: *filter photometers* and *spectrometers*. Two factors that distinguish UV instruments from visible light instruments are the need for UV transmitting optical components and the need to avoid stray light from the visible solar spectrum. Ordinary glass absorbs light at wavelengths less than 400 nm, so materials like quartz are required for transmitting optical components. The UV tail of the solar spectrum contains less than 10 % of the energy, and absorption by ozone can reduce radiances by three orders of magnitude lower than the solar irradiance. Care must be taken to insure that no visible light is transmitted or scattered in the optical system to be detected along with the desired UV wavelength. This becomes increasingly important as the band pass decreases or as the selected wavelengths are absorbed by the atmosphere. Stray light rejection is usually accomplished by use of filters, such as NiSO₄ crystals, or dichroic mirrors or by use of double monochromators. UV instruments are described in books by Huffman (1992) and Webb (1998).

Ground-based remote sensing

Atmospheric ozone absorbs ultraviolet sunlight at wavelengths less than 330 nm, increasing to a maximum near 250 nm. Dobson and Harrison (1926) developed a spectrometer to measure the total column amount of ozone in one of the first UV remote sensing experiments. They used a double-beam, double-prism spectrometer to measure the wavelength differential absorption by ozone in direct sunlight while removing visible light from the detector. Three pairs of wavelengths were needed for measurements over the full range of ozone amounts and solar zenith angles. A wedge neutral density filter is moved across the brighter beam until the intensities are matched as detected by a photomultiplier. An extraterrestrial calibration is derived from a Langley plot of measurements at each wavelength pair over a wide range of solar zenith angles on a day with constant total ozone. The changing path is extrapolated to zero to simulate the extraterrestrial reading. Dobson spectrophotometers are still widely used for monitoring of total ozone (Dobson, 1968). Vertical ozone profile information was also derived from zenith sky observations during twilight using an Umkehr (reversal) technique. Later, Brewer (1973) produced a grating spectrometer to measure total ozone. This instrument and a double grating version are also in wide use around the world.

Rockets

The first successful measurement of the vertical ozone distribution above 30 km used a rocket-borne photographic UV spectrograph to measure the extinction of sunlight as a function of altitude (Johnson et al., 1952). The ozone density decreases exponentially with altitude and a set of wavelengths with successively higher ozone absorption coefficients was needed to maintain precision with altitude. Following this pioneering measurement, a low-cost UV filter photometer ozonesonde was developed for launch from many global stations on meteorological sounding rockets (Krueger, 1973). This instrument used three interference filters and a nickel sulfate – Corning glass filter combination to block visible light. The data were taken during descent on a parachute from 60 km, where the extraterrestrial solar intensity was measured for computation of overhead ozone at lower altitudes. Sunlight was incident on a flat diffuser plate and a beam splitter corrected signal variations due to angle changes. The US Standard Atmosphere (1976) ozone distribution was constructed from rocket sounding data.

Satellite UV instruments

When satellites were developed, ultraviolet instruments were an early priority because the atmosphere absorbs these wavelengths. Initially, filter photometers using the same components as rocketsondes were flown. A nadir-viewing instrument (Friedman et al., 1963), launched in 1962, measured radiances at 255 nm for comparison with early radiative transfer calculations (Green, 1966). A 260 nm filter UV radiometer was also

flown in 1962 to measure the vertical ozone profile from the extinction of sunlight during sunrise and sunset (Rawcliffe et al., 1963). Flights of UV spectrometers began 3 years later.

Spectrometers

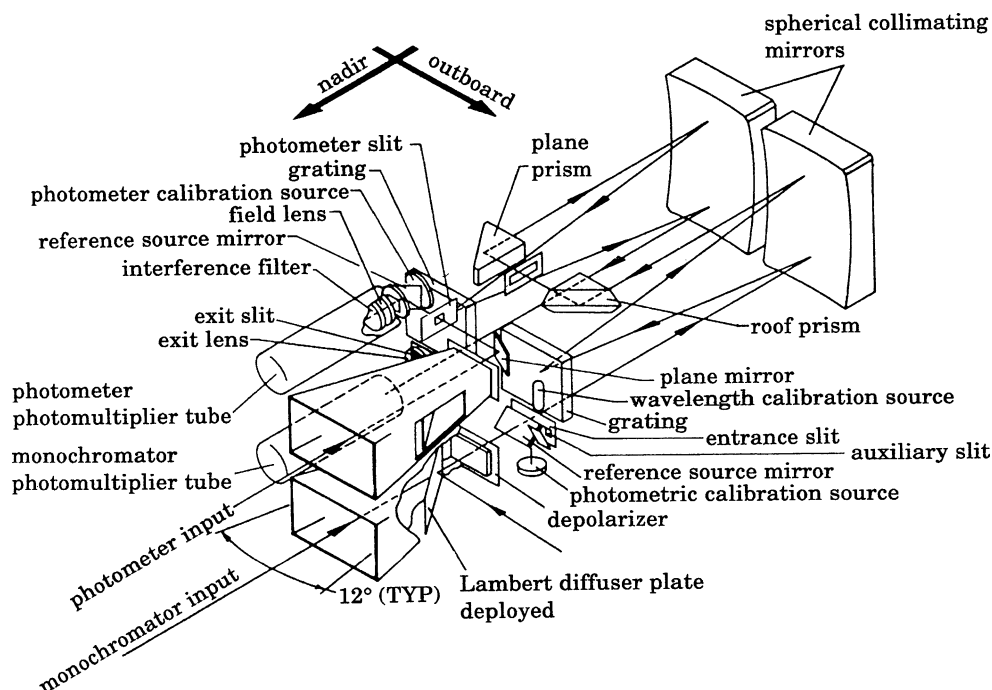
Iozenas et al. (1969) used a double monochromator with a spectral resolution of 1.4 nm on satellite flights in April 1965 and June 1966 to measure nadir radiance and solar irradiance from which the atmospheric spectral albedo was computed. The solar spectra were determined using a matte quartz diffuser plate to fill the instrument field of view (FOV). Limb spectra were also measured using a mirror to point the instrument axis slightly above the visible limb. In 1967, C. A. Barth launched a single Ebert-Fastie monochromator on the NASA OGO-4 satellite. Anderson et al. (1969) used a single-scattering ozone retrieval algorithm (Twomey, 1961) and calibrated the albedos using a coincident rocket sounding (Krueger, 1969).

BUV/SBUV

The first satellite instrument optimally designed for global ozone measurements was the NASA Backscatter Ultraviolet (BUV) instrument that was proposed by C.L. Mateer and J.V. Dave of NCAR and developed by D.F. Heath of Goddard Space Flight Center. Mateer's retrieval algorithm and Dave's radiative transfer calculations were used for vertical profile retrievals and also to test an algorithm for total ozone retrievals (Dave and Mateer, 1967). The BUV and its successor, the Solar Backscatter Ultraviolet (SBUV) instrument on Nimbus-7, used eight selected wavelengths between 255 and 308 nm for profile retrievals and four wavelengths between 312.5 and 339.8 nm for total ozone retrievals. Major design issues were the three orders of magnitude change in the albedo between 250 and 340 nm due to ozone absorption and stray light rejection requirement of six orders of magnitude. The BUV/SBUV is a cam-driven, wavelength step scan, double Ebert-Fastie monochromator (Figure 1) with very high stray light rejection and a large dynamic range (Heath et al., 1973). The wavelength scanning was slow (32 s/frame) for good signal/noise ratio at the short wavelengths, and a co-aligned 380 nm filter photometer channel measured radiance changes due to underlying clouds in the nadir-pointed FOV as the spacecraft moved in its orbit. The instrument response to solar irradiance was measured using a diffuser plate that was moved into the FOV at the terminator. Degradation of the reflectivity of the diffuser plate affected the albedo calibration. This became a significant issue when ozone trends were sought to confirm predictions of ozone depletion. New, redundant diffuser designs were incorporated in later flights.

TOMS

The total ozone retrieval scheme of Dave and Mateer (1967), tested on the Nimbus-4 BUV, delivered results in



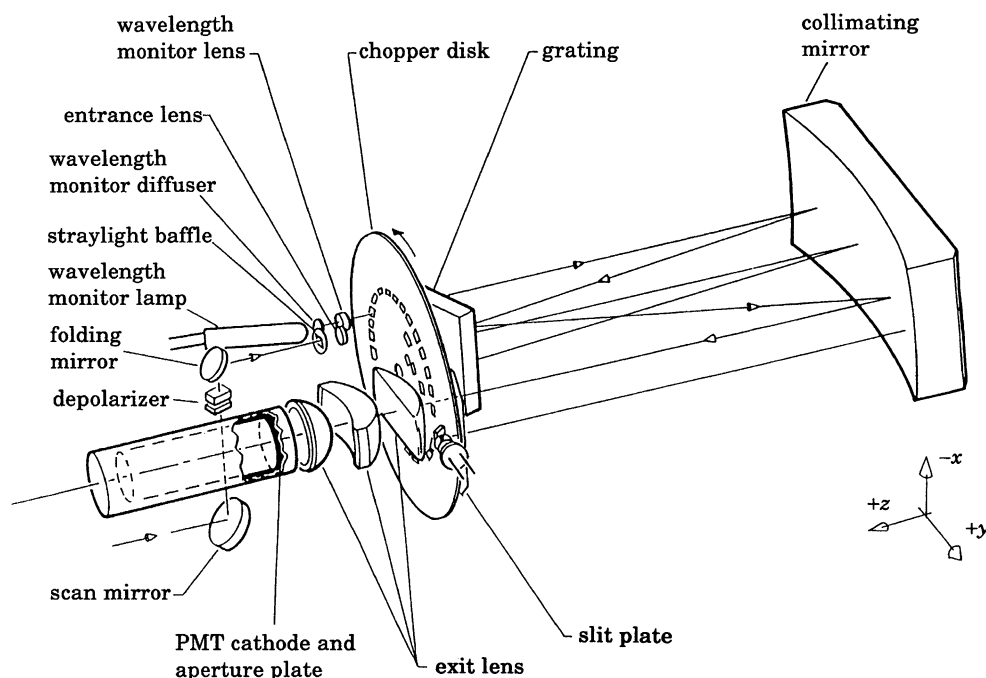
Ultraviolet Sensors, Figure 1 Optics diagram of the BUV and SBUV instruments. BUV-class instruments are double Ebert-Fastie monochromators that accept backscattered sunlight from an 11° nadir field of view at 12 selected 1 nm wavelength channels between 255 and 340 nm. A depolarizer negates any polarization sensitivity. The wavelengths are selected serially over a 32 s cycle using cam-driven holographic gratings. The radiance is measured with a photomultiplier detector. The Earth albedo calibration is obtained from direct solar irradiance data taken via a diffuser plate that is rotated into the field of view when the Sun is near the horizon at sunset. Fluctuations in radiance due to changing cloud fields are compensated using a parallel filter photometer channel that is frequently sampled.

excellent agreement with coincident Dobson spectrometer data. However, the high variability suggested structure in the total ozone field. A. J. Krueger of Goddard Space Flight Center proposed a new instrument, the Total Ozone Mapping Spectrometer (TOMS), to map this structure by scanning the FOV cross track while measuring the high-radiance, total ozone wavelengths to permit short sampling times. The TOMS instrument (Figure 2) uses a single Ebert-Fastie polychromator with six, fixed exit slits corresponding to the BUV total ozone wavelengths and two longer wavelengths to measure surface spectral reflectivity. A chopper wheel sequentially passes the wavelengths to a photomultiplier. A scan mirror synchronously directs the FOV to 35 cross-track positions. A wavelength scrambling depolarizer avoids biases due to monochromator polarization. Scanning is designed for complete global coverage each day with 50×50 km nadir resolution that is adequate to resolve suspected jet stream modulations on the total ozone field yet did not overload the Nimbus telemetry. The TOMS instrument was mated with an advanced BUV, the Solar Backscatter Ultraviolet (SBUV) instrument, for flight on the Nimbus-7 satellite in 1978 (Heath et al., 1975).

GOME

European interest in satellite-based UV remote sensing rose in the 1990s as ozone became an environmental issue. John Burrows of the Max Planck Institute/Mainz proposed a new instrument for the ERS-2 satellite that took advantage of many improvements in satellite technology since 1970. In addition, new solid-state UV detector arrays were available. The Global Ozone Monitoring Instrument (GOME) measures the full spectrum from 240 to 790 nm rather than discrete wavelengths as used in the BUV class of instruments. The spectrum is first split into four channels using quartz pre-disperser prisms (Figure 3). Each channel then is dispersed by a holographic grating to a Reticon linear array silicon detector. To correct for the polarization sensitivity of the instrument, a device is added to measure the degree of polarization in three broadband channels. The instrument field of view (40×320 km) is directed to three cross-track positions as the spacecraft flies in a polar orbit.

The GOME instrument design is tailored for use of the Differential Optical Absorption Spectroscopy (DOAS) algorithm for retrieval of total ozone and other NUV absorbing species. This differs from the SBUV and TOMS



Ultraviolet Sensors, Figure 2 Optics diagram of TOMS instruments. TOMS instruments consist of single Ebert-Fastie polychromators that accept backscattered sunlight from a 3° field of view that is directed in 35 cross-track positions with a scanning mirror. The swath width of 2,700 km provides contiguous global coverage each day with 50 km nadir ground resolution. Total ozone and sulfur dioxide and aerosol index are derived from the radiance at six selected 1 nm wavelength channels between 312.5 and 380 nm on the Nimbus-7 and Meteor-3/TOMS missions. The wavelengths are sampled using a rotating aperture plate that opens the six exit slits sequentially. A depolarizer after the scan mirror avoids polarization sensitivity biases. The Nimbus TOMS used the SBUV diffuser plate for solar calibration, while later instruments used a carousel to rotate three diffuser plates into the field of view using different exposure rates to correct for degradation.

that were designed to use strongly absorbing wavelength bands and radiative transfer tables to retrieve ozone amounts. The DOAS method makes use of spectral band fine structure at weakly absorbed, longer wavelengths and is less dependent on absolute radiance calibrations than is the BUW method. With the continuous spectral coverage of GOME, trace gases including formaldehyde and BrO, in addition to NO_2 which absorbs at violet wavelengths, were measured for the first time (Burrows et al., 1999).

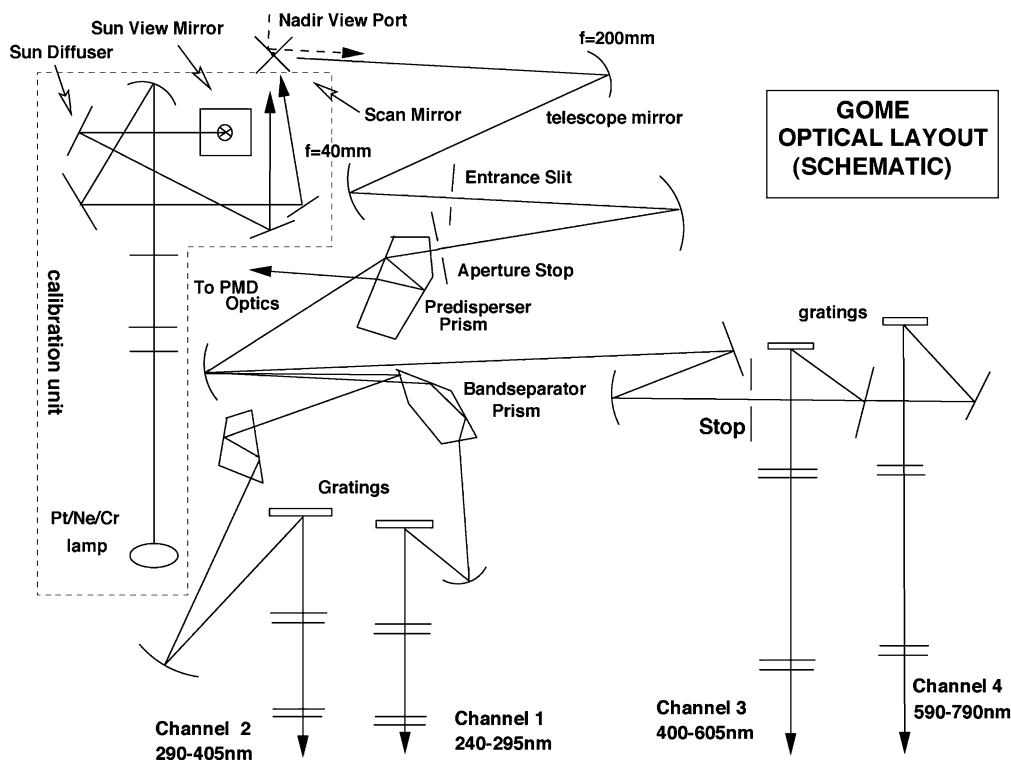
SCIAMACHY

A new Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) research instrument was launched on the European environmental satellite ENVISAT in 2002. SCIAMACHY alternated nadir and limb observations for more information on the profiles of stratospheric constituents using scattered light at the horizon. Nadir spatial resolution was improved over GOME by scanning the FOV across the flight track with a ground resolution of 60×30 km and a swath width of 960 km. An online book, SCIAMACHY: Monitoring the Changing Earth's Atmosphere, is available at: <http://atmos.caf.dlr.de/projects/scops/>

sciamachy_book/sciamachy_book.html. Many of the scientific products are available through <http://www.iup.uni-bremen.de/sciamachy/index.html>.

OMI

The next advancement in UV Earth observations was the Ozone Monitoring Instrument (OMI) built by the Netherlands and Finland for flight on the NASA EOS Aura Platform in 2004 (Levelt et al., 2006). OMI is a complex hyperspectral grating spectrometer (Figure 4) using a thin (1°) along-track but wide (114°) cross-track field of view in a push broom configuration to obtain full contiguous daily global observations at UV (264–380 nm) and visible (349–504 nm) wavelengths (<http://www.knmi.nl/omi/research/instrument/index.php>). A depolarizer avoids radiometric biases due to polarization sensitivity in the spectrometer. A dichroic mirror directs visible and UV wavelengths into separate optical channels for stray light control. The UV light then is divided into short (264–311 nm) and long (307–383 nm) wavelength regions again to minimize stray light at the short, highly absorbed wavelengths used for ozone profile retrievals. The dispersed light is imaged on a frame-transfer back-thinned CCD with cross-track and spectral information



Ultraviolet Sensors, Figure 3 Optics diagram of the GOME instrument. GOME is a double UV/visible spectrometer with continuous spectral coverage from 240 to 790 nm with moderate resolution. The full spectrum is split into four channels by a group of prisms and a beam splitter. Each channel is then dispersed with a grating and detected with a silicon linear array with 1,024 pixels. A portion of the light is sent to a polarization-measuring device (PMD) for correction of the radiance measurements. Radiance data are collected over three large (40×320 km) ground pixels.

as orthogonal components. Thus, as the spacecraft moves along its orbit, the spatial-spectral data are serially read out for telemetry to the ground. An eightfold improvement in ground resolution over TOMS enhances detection of small sources of SO_2 and absorbing aerosols. This instrument returned to the TOMS concept of contiguous daily global coverage but with an eightfold increase in ground resolution (13×24 km at nadir) and with full spectral coverage. The contiguous coverage has resulted in exceptional daily images of NO_2 and SO_2 sources. The availability of optimum wavelengths results in a 20-fold improvement in SO_2 sensitivity over TOMS, allowing monitoring of volcanic degassing, smelter pollution, and pollution from fossil fuel burning.

GOME-2

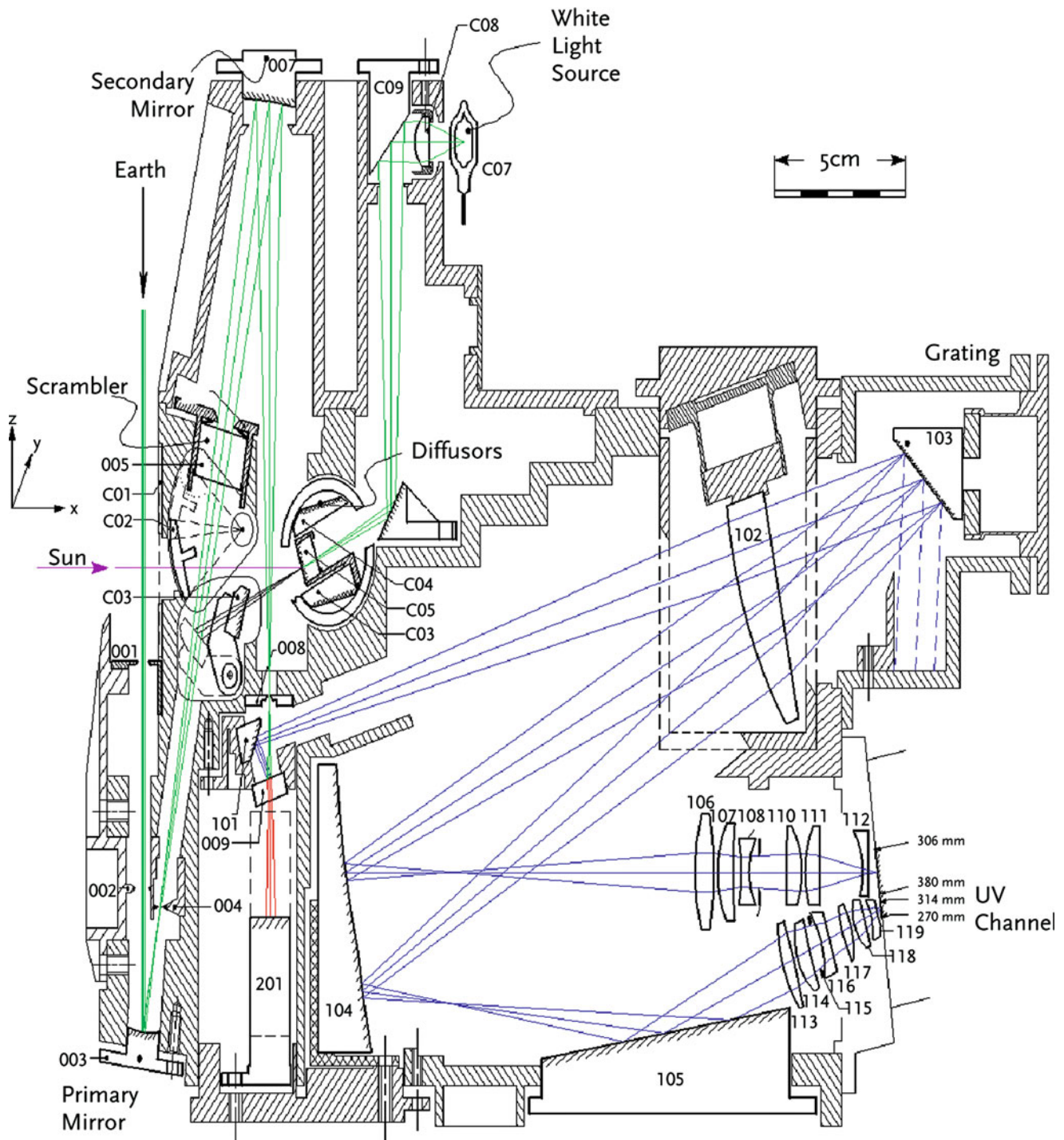
A second-generation GOME instrument, GOME-2, was launched on the first European polar-orbiting operational meteorological satellite, Metop-A, in 2006 (<http://www.eumetsat.int/Home/Main/Satellites/Metop/index.htm>). This was followed in 2012 with a second GOME-2 instrument on the Metop-B satellite. GOME-2 adds a spatial scanner with ground resolution similar to TOMS. The series of Metop flights will provide a continuation of the long-term record of ozone trends.

The GOME-2 instrument and mission are described in http://www.esa.int/esaLP/SEMTEG23IE_LPmetop_0.html.

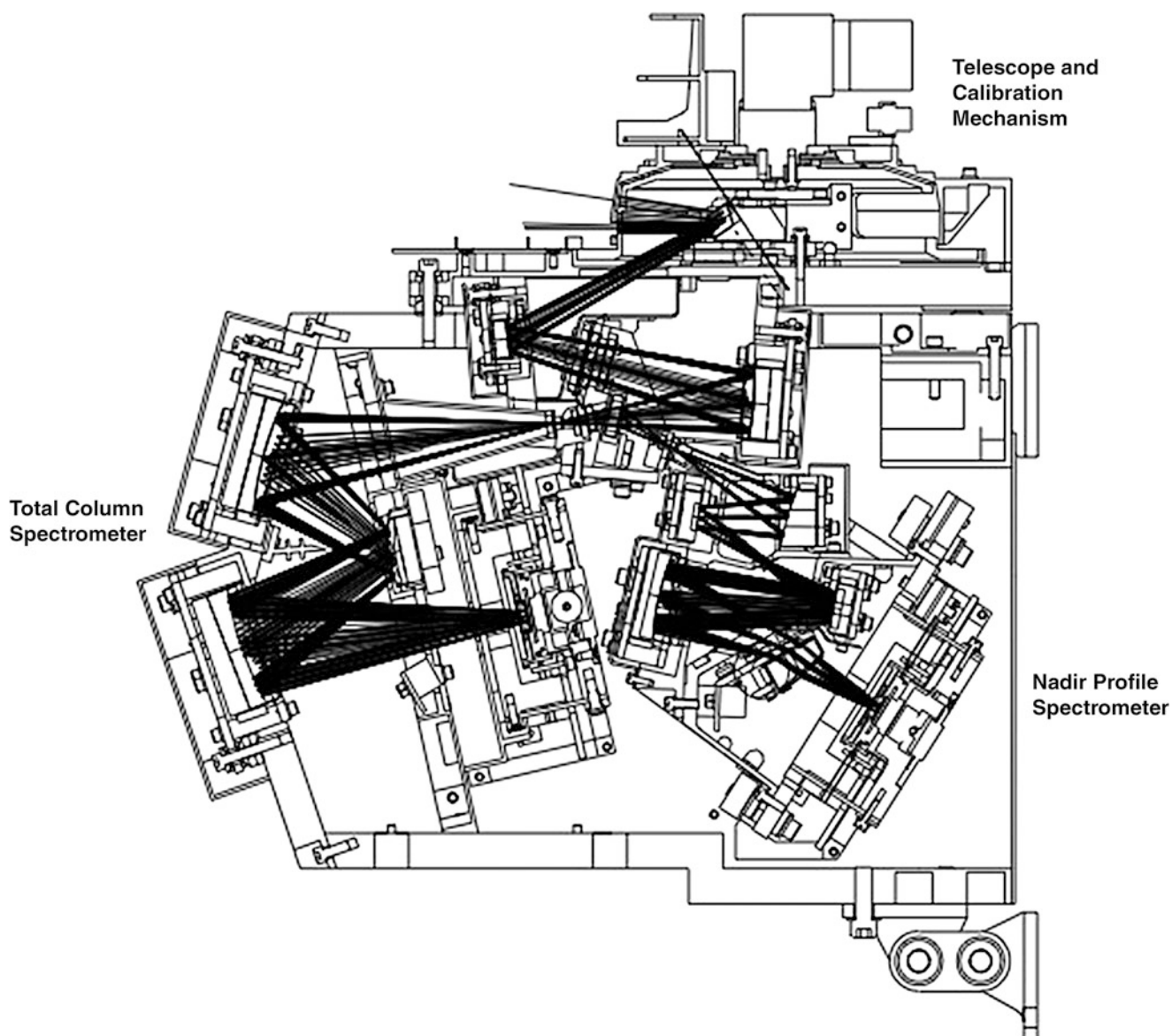
OMPS

The first American hyperspectral UV remote ozone sensing instrument, Ozone Mapping and Profiler Suite (OMPS), was launched in 2011 on the Suomi National Polar-orbiting Partnership (NPP) satellite (<http://npp.gsfc.nasa.gov/omps.html>). OMPS consists of a nadir sensor with requirements similar to TOMS and SBUV and a limb imaging spectrometer for ozone profiling with higher vertical resolution than is possible with nadir sounding. The NPP OMPS is the first of a series planned for Joint Polar Satellite System (JPSS) operational missions. OMPS continues ozone trend monitoring that began with the Nimbus-7 SBUV/TOMS mission in 1978.

The OMPS Nadir Sensor (Dittman et al., 2002a) consists of total column and profile ozone spectrometer modules with a common telescope operated in a wide field-of-view (FOV) push broom configuration. A dichroic beam splitter divides the 250–380 nm Earth radiance spectrum near 305 nm into the longer, total ozone and the shorter, ozone profile wavelengths. A blocking filter reduces visible wavelength radiances for stray light control. For total ozone retrievals the 300–380 nm



Ultraviolet Sensors, Figure 4 Optics diagram of the OMI instrument. OMI uses a dichroic mirror to separate the spectrum into visible and UV components. This diagram shows the UV section where NUV and MUV wavelengths are dispersed by a grating after collimation by a silica lens. The dispersed light is focused near a field mirror that splits the spectrum into NUV and MUV components for stray light control. Both UV spectra are imaged on a frame-transfer CCD detector. The instrument field of view is a $114^\circ \times 1^\circ$ fan produced by an entrance telescope. The cross-track information is focused in one CCD dimension, while spectral information is in the other dimension. A depolarizer corrects any polarization sensitivity. The ground resolution is 13×24 km and the swath width produces full contiguous daily global resolution.

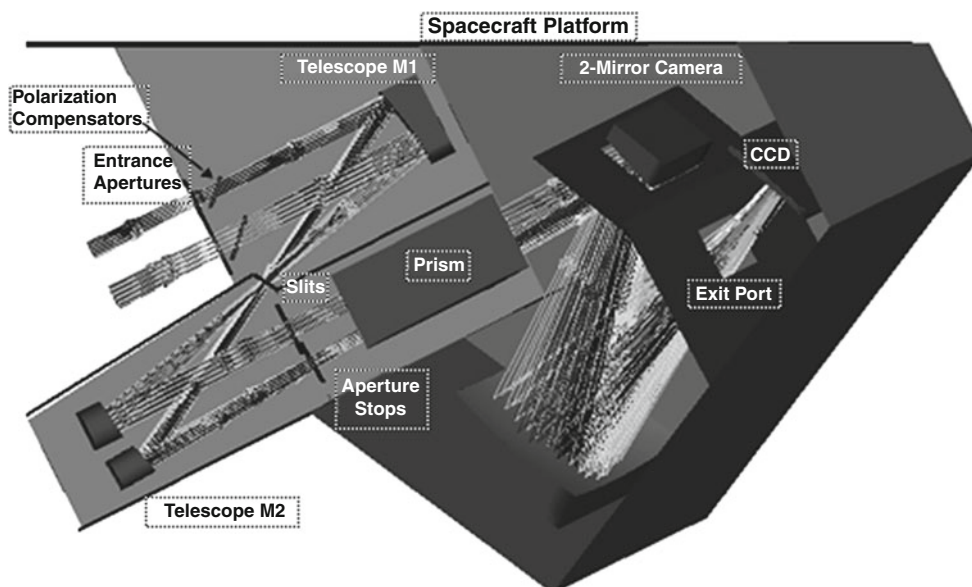


Ultraviolet Sensors, Figure 5 Optical configuration of the OMPS Nadir Sensor – a depolarized push broom hyperspectral UV imaging instrument with contiguous global coverage. A total column ozone module uses a single grating spectrometer for the longer, higher radiance total ozone wavelengths (300–380 nm); an ozone profile module uses double gratings for the shorter, low-radiance wavelengths (250–310 nm). Both spectrometers share a push broom telescope with a 110° cross-track field of view for contiguous global total ozone coverage.

backscattered radiance across a 110° FOV is dispersed by a convex grating spectrometer and imaged on a split frame-transfer, cooled CCD. For the 250–310 nm ozone profile wavelengths, the radiance from a 16.7° central part of the FOV is dispersed in a Monk-Gillieson double spectrometer and imaged on a second CCD. Spectral resolution is 1 nm. Solar irradiance is measured with working and reference diffusers for degradation assessment (Figure 5).

The OMPS Limb Imaging Spectrometer (Dittman et al., 2002b; Leitch et al., 2003) measures the radiance

distribution in vertical traces across the Earth's daytime limb to retrieve vertical ozone and aerosol distributions with 3 km vertical resolution. Upper stratospheric ozone distributions to 60 km are derived from 10 selected UV wavelengths; aerosol profiles and lower stratospheric ozone distributions above 8 km are derived from 11 visible and near-infrared wavelengths. A prism spectrometer is fed by 3 telescopes with horizontally displaced views of the limb to estimate horizontal gradients in ozone. Height versus wavelength radiances from three limb scans are simultaneously imaged on a split frame-transfer CCD.



Ultraviolet Sensors, Figure 6 Principal components of the OMPS Limb Imaging Spectrometer. Each of three telescopes utilizes confocal pairs of off-axis parabolas (M1 and M2). Dual entrance apertures (for dynamic range) are followed by a polarization compensator, entrance slits, and aperture stops. The secondary mirrors M2 feed a quartz prism spectrometer. A two-mirror camera focuses the images on a CCD.

The five order-of-magnitude change of radiances across the limb and the spectrum is addressed by careful stray light control, dual aperture stops, and multiple gain ranges. Pixel gain on orbit is monitored using solar-illuminated working and reference diffuser plates (Figure 6).

Far and extreme UV sensors

A long series of experiments have been flown to measure UV airglow and auroral emissions from space. The first FUV image of the Earth was taken from the Moon during the Apollo program (Carruthers and Page, 1972), and Gérard and Barth (1976) measured auroral spectra from OGO-4. Auroral images at H Lyman alpha (121.6 nm) were taken with a modified TV camera on the Japanese EXOS – K satellite and University of Iowa spin scan filter photometers on Dynamics Explorer satellites produced images of the aurora in the 1980s (Frank and Craven, 1988). Several satellite programs followed, including High LATitude (HILAT) and (Polar Beacon experiment and Auroral Research) Polar BEAR, carrying photometers and Ebert-Fastie spectrometers. Most of the geophysical information is in the 100–200 nm FUV spectral region. The NASA POLAR mission (<http://pwg.gsfc.nasa.gov/polar/>), launched in 1996, includes an Ultraviolet Imager to detect auroral emissions in the 130–190 nm region as part of the International Solar Terrestrial Physics Program. The TIMED mission, launched in 2001, includes a Global Ultraviolet Imager (115–180 nm) in a group of instruments to investigate the mesosphere and lower thermosphere (<http://www.timed.jhuapl.edu/WWW/index.php>). The cited web sites are excellent sources of information

on these missions, which are designed to support Space Weather research.

Summary

UV remote sensing instruments are used to measure atmospheric ozone at near and middle UV wavelengths and to measure airglow and auroral emissions at far UV wavelengths. Ozone measurements made from the ground, from balloons and rockets, and from space use the absorption of sunlight in Hartley and Huggins bands of ozone. Total column amounts are measured from the ground using monochromators. The vertical distribution of ozone is measured with interference filter photometers on rockets and balloons. Both total column and vertical ozone distributions are remotely measured from satellites using filter photometers and spectrometers that view the sunlit nadir atmosphere or the limb.

In the USA, the Backscatter Ultraviolet (BUV) instrument was launched on the Nimbus-4 satellite in 1970. SBUV and Total Ozone Mapping Spectrometer (TOMS) instruments flown on Nimbus-7 in 1978 and on later satellites have collected the longest satellite record of ozone changes. Both instruments use Ebert-Fastie monochromators. The BUV and successor, Solar Backscatter Ultraviolet (SBUV), instruments require double monochromators to reject stray light from visible wavelengths that dominate the solar spectrum. These instruments measure the Earth radiance at selected wavelengths and use depolarizers to avoid sensitivity to polarization in Rayleigh-scattered sunlight. They are calibrated by viewing a diffuser plate exposed to extraterrestrial sunlight.

Filter photometers and Ebert-Fastie monochromators are also used at far UV wavelengths to measure spectral line emissions from the thermosphere and ionosphere.

New generations of European ozone monitoring instruments make use of composite optical benches for full spectrum data at UV and visible wavelengths. The first instrument, Global Ozone Monitoring Instrument (GOME), launched on ERS-2 in 1995, uses a pre-disperser prism and gratings to break the spectrum into four bands that individually measure the nadir spectral radiance with linear detector arrays. A successor, the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY), launched on ENVISAT in 2002, adds limb data and better ground resolution. Full daily global coverage, similar to TOMS, was obtained with the Dutch-Finnish Ozone Measuring Instrument (OMI), launched on the EOS Aura satellite in 2004. This instrument avoids a mechanical scanner by using a fan-shaped field of view operated in a push broom fashion. OMI uses a dichroic mirror to separate the spectrum into bands that are then dispersed with gratings for detection with a CCD containing orthogonal spatial and spectral information. European operational ozone monitoring is carried out with a new generation series of GOME instruments, GOME-2, first launched on the Metop-A satellite in 2006, next on the Metop-B satellite in 2012. This instrument adds spatial scanning similar to TOMS. In the USA, operational ozone monitoring is being continued using the Ozone Mapping and Profile Suite (OMPS) series of hyperspectral instruments, initially launched on the Suomi – NPP satellite – in 2011, to be followed by flights on JPSS satellites.

Bibliography

- Anderson, G. P., Barth, C. A., Cayla, F., and London, J., 1969. Satellite observations of the vertical ozone distribution in the upper stratosphere. *Annals of Geophysics*, **25**, 341–345.
- Brewer, A. W., 1973. A replacement for the Dobson spectrophotometer. *Pure and Applied Geophysics*, **106–108**, 919–927.
- Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V. V., Ladstädter-Weissenmayer, A., Richter, A., de Beek, R., Hoogen, R., Bramstedt, K.-U., Eichmann, K., Eisinger, M., and Perner, D., 1999. The global ozone monitoring experiment (GOME): mission concept and first scientific results. *Journal of the Atmospheric Sciences*, **56**, 151–175.
- Carruthers, G. R., and Page, T., 1972. Apollo 16 far ultraviolet camera spectrograph: earth observations. *Science*, **177**, 788–791.
- Dave, J. V., and Mateer, C. L., 1967. A preliminary study on the possibility of estimating total atmospheric ozone from satellite measurements. *Journal of the Atmospheric Sciences*, **24**, 414–427.
- Dittman, M. G., Ramberg, E., Chrisp, M., Rodriguez, J. V., Sparks, A. L., Zaun, N. H., Hendershot, P., Dixon, T., Philbrick, R. H., and Wasinger, D., 2002a. Nadir ultraviolet imaging spectrometer for the NPOESS Ozone Mapping and Profiler Suite (OMPS). In *SPIE Proceedings*, Vol. 4814. Earth Observing Systems VII, William L. Barnes, Editor, pp. 111–119.
- Dittman, M. G., Leitch, J. W., Chrisp, M., Rodriguez, J. V., Sparks, A. L., McComas, B. K., Zaun, N. H., Frazier, D., Dixon, T., Philbrick, R. H., Wasinger, D., 2002b. Limb broad-band imaging spectrometer for the NPOESS Ozone Mapping and Profiler Suite (OMPS). In *SPIE Proceedings*, Vol. 4814. Earth Observing Systems VII, William L. Barnes, Editor, pp. 120–130.
- Dobson, G. M. B., 1968. Forty years' research on atmospheric ozone at Oxford: a history. *Applied Optics*, **7**, 387–405.
- Dobson, G. M. B., and Harrison, D. N., 1926. Measurements of the amount of ozone in the earth's atmosphere and its relation to other geophysical conditions. *Proceedings of the Royal Society London A*, **1926**(110), 660–69.
- Frank, L. A., and Craven, J. D., 1988. Imaging results from dynamics explorer I. *Reviews of Geophysics*, **26**, 249–283.
- Friedman, R. M., Rawcliffe, R. D., and Meloy, G. E., 1963. Radiance of the upper atmosphere in the middle ultraviolet. *Journal of Geophysical Research*, **68**, 6419–6423.
- Gérard, J.-C., and Barth, C. A., 1976. OGO-4 observations of the ultraviolet auroral spectrum. *Planetary and Space Science*, **24**, 1059–1063.
- Green, A. E. S., 1966. *The Middle Ultraviolet: Its Science and Technology*. New York: Wiley.
- Heath, D. F., Mateer, C. L., and Krueger, A. J., 1973. The Nimbus 4 Backscattered Ultraviolet (BUV) atmospheric ozone experiment – two years operation. *Pure and Applied Geophysics*, **106–108**, 1238–1253.
- Heath, D. F., Krueger, A. J., Roeder, H. A., and Henderson, B. D., 1975. The solar backscatter ultraviolet and total ozone mapping spectrometer (SBUV/TOMS) for Nimbus G. *Optical Engineering*, **14**, 323–331.
- Huffman, R. E., 1992. *Atmospheric Ultraviolet Remote Sensing*. San Diego: Academic.
- Iozenas, V. A., Krasnopol'skiy, V. A., Kuznetsov, A. P., and Lebedinsky, A. I., 1969. Studies of the Earth's ozonosphere from satellites. *Atmospheric and Oceanic Physics*, **5**, 149–159.
- Johnson, F. S., Purcell, J. D., Tousey, R., and Watanabe, K., 1952. Direct measurements of the vertical distribution of atmospheric ozone to 70 kilometers altitude. *Journal of Geophysical Research*, **57**, 157–176.
- Krueger, A. J., 1969. Rocket measurements of ozone over Hawaii. *Annals of Geophysics*, **25**, 307–311.
- Krueger, A. J., 1973. The mean ozone distribution from several series of rocket soundings to 52 km at latitudes from 58° S to 64° N. *Pure and Applied Geophysics*, **106**(1), 1272–1280.
- Leitch, J. W., Rodriguez, J. V., Dittman, M. G., Frazier, D., McComas, B. K., Philbrick, R. H., Wasinger, D., Valle, T. J., Dixon, T., Dooley, D., Munzer, R., and Larsen, J. C., 2003. Limb broad-band imaging spectrometer for the NPOESS Ozone Mapping and Profiler Suite (OMPS). In *SPIE Proceedings*, Vol. 4891, pp. 13–21.
- Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J., and Saari, H., 2006. The ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(5), 1093–1101. 10.1109/TGRS.2006.872333.
- Rawcliffe, R. D., Meloy, G. E., Friedman, R. M., and Rogers, E. H., 1963. Measurement of vertical distribution of ozone from a polar orbiting satellite. *Journal of Geophysical Research*, **68**, 6425–6429.
- Twomey, S., 1961. The deduction of the vertical distribution of ozone by ultraviolet spectral measurements from a satellite. *Journal of Geophysical Research*, **66**, 2153–2162.
- U.S. Standard Atmosphere, 1976. U.S. Government Printing Office, Washington, D.C., 1976.
- Webb, A. R., 1998. *UVB Instrumentation and Applications*. Amsterdam: Gordon and Breach Science.

Cross-references

[Air Pollution](#)
[Climate Monitoring and Prediction](#)
[Geophysical Retrieval, Overview](#)
[Stratospheric Ozone](#)

Trace Gases, Stratosphere, and Mesosphere
 Trace Gases, Troposphere - Detection from Space
 Ultraviolet Remote Sensing
 Volcanism

URBAN ENVIRONMENTS, BEIJING CASE STUDY

Son V. Nghiem¹, Alessandro Sorichetta²,
 Christopher D. Elvidge³, Christopher Small⁴,
 Deborah Balk⁵, Uwe Deichmann⁶ and Gregory
 Neumann¹

¹Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

²Dipartimento di Scienze della Terra “A. Desio”,
 Università degli Studi di Milano, Milan, Italy

³Earth Observation Group, NOAA-NESDIS National
 Geophysical Data Center E/GC2, Boulder, CO, USA

⁴Lamont Doherty Earth Observatory, Marine Geology and
 Geophysics, Columbia University, Palisades, NY, USA

⁵School of Public Affairs, Baruch College, City
 University of New York, New York, NY, USA

⁶Development Research Group, The World Bank,
 Washington, DC, USA

Definition and key parameters

Remote sensing data are used together with demographic parameters to characterize and monitor **urban environments**. For urban remote sensing, key measurements include night-light (NL) from visible-near infrared sensors, multispectral signatures having different spectral mixing (SM) for different land cover categories, and high-resolution radar backscatter (σ_0) obtained from scatterometer data with the Dense Sampling Method (DSM). NL is defined as the light occurred in areas lit by anthropogenic sources at nighttimes. SM is a composition of signatures from multiple channels with different electromagnetic wavelengths corresponding to blue, green, and red in the visible light range and also to parts of the near infrared. DSM σ_0 in an urban area represents the radar cross section from ensembles of targets such as houses, buildings, factories, and other man-made infrastructures on land surface. Associated with σ_0 for each target ensemble, an index of variability (IV) is defined as $(1 + \delta/\sigma_0)$ in decibel, where δ is the standard deviation. Among the critical datasets to describe social and economic patterns are population and road network. The residential population (P_r) is defined as the population counted in each administrative unit. The ambient population (P_a) represents the number of daily averaged people located in each gridded pixel. Road network data are available in many formats from various sources that are routinely updated. These parameters will be further described in detail in the sections below. In this entry, Beijing in China is used to illustrate the general points.

Urban landscape

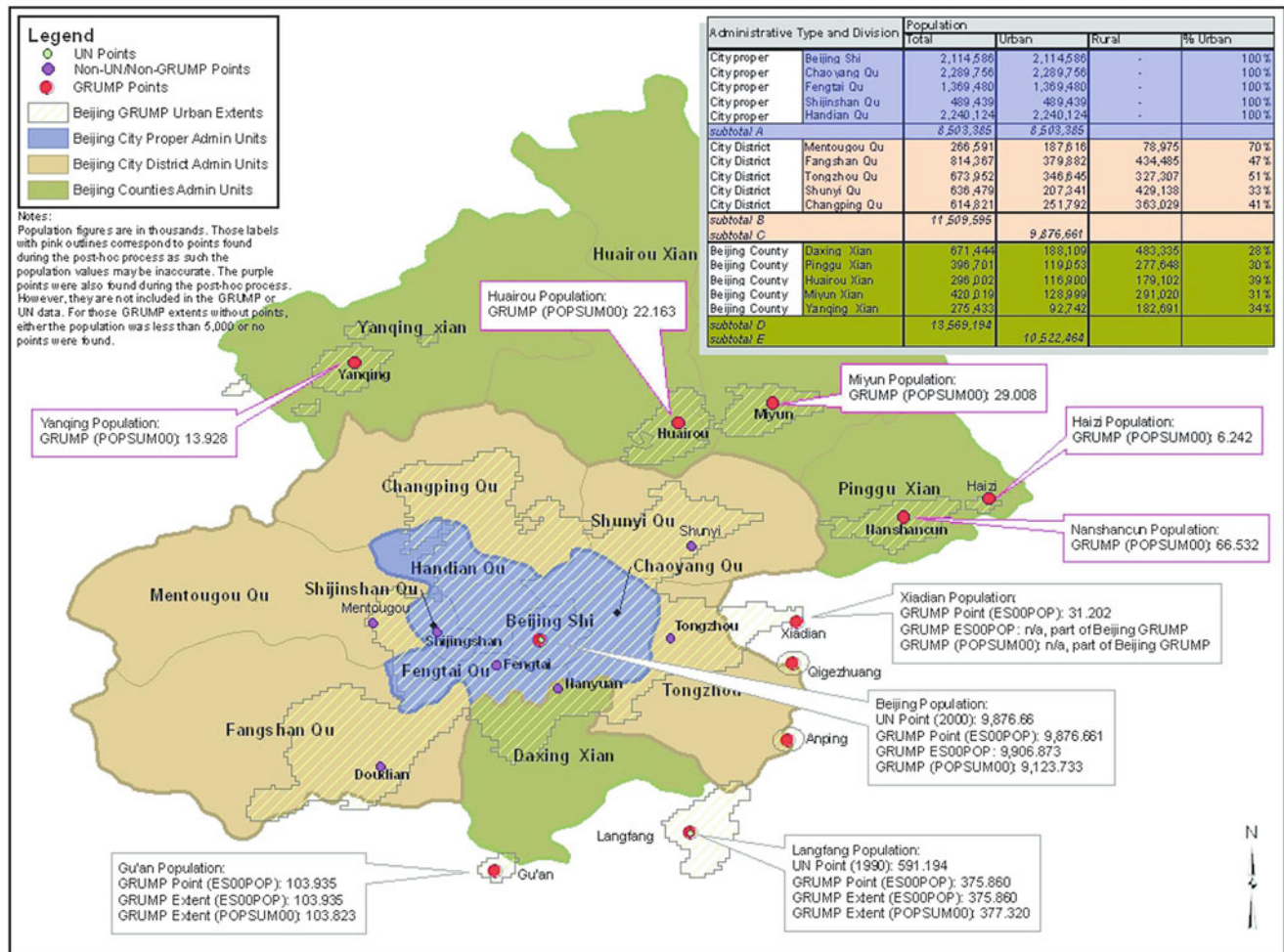
Remote sensing technology offers some unique images and data of the urban landscape. Much of this entry explains the contributions that remotely sensed data make independent of other data sets. In this section, we also describe the contribution of such data when used in combination with demographic and socioeconomic data. Particularly where the latter types of data are lacking, the combination of data sources may be especially powerful for identifying and monitoring the physical and human urban landscape.

Cities exist largely because people and capital become more productive when they cluster close together. Yet, for urban growth to be beneficial, urban policy making must reconcile three main objectives: economic growth, equity among the population, and environmental sustainability. Additionally, to achieve these objectives, policy makers must facilitate mobility which requires setting aside sufficient land for public and private forms of transport. Maintaining environmental sustainability also requires good institutions and enforcement for making smart land-use decisions that encourage efficient transport and a safe built environment. Moreover, climate change and potential exacerbation of disasters such as flood, drought, and fire extremes are major factors in long-term sustainable urban planning.

Designing the right policy instruments to help make urban growth productive and sustainable requires up-to-date information. However, most cities do not have the capacity to maintain up-to-date land registries and land-use maps at a fine scale. Population and housing censuses provide some information about who lives where and the characteristics of the dwellings and residents, but are typically collected only once per decade. Information on individual parcels of land (e.g., through tax records) is even sparser in most countries. Even where such statistical systems are introduced, they lack a retrospective perspective. Remote sensing data can help fill these data gaps and improve urban management, planning, and monitoring.

Here, Beijing is considered as a case study. This city represents an example of rapid urban growth in an emerging economy. Since the 1980s, China implemented containment strategies to guide urban growth. In the face of double-digit economic growth concentrated in the urban areas in Eastern China, Beijing and other cities have nevertheless struggled to address the needs for economic growth, rapid motorization driven by rapid income growth, huge demand for housing from migrants, and wealth-driven demand for larger floor space. Between 1990 and 2008 its population grew from 10.8 to 17 million on a land area of 16,410 km². Population is coarsely allocated into administrative units (Figure 1), which are arbitrarily defined with boundaries determined by obscure political processes. Further compounding the uncertainty, these are official figures and do not include the “floating” population – migrants who do not hold official, *Hukou*, registration cards (Chengrong et al., 2008; Yeh and Xu, 2009). Because the

Beijing Administrative Units and Urban Allocation



Urban Environments, Beijing Case Study, Figure 1 Beijing administrative unit and urban allocation. NL9495 extent is denoted by the hatched areas. Blue, brown, and green areas indicate the administrative areas for Beijing city proper, city district, and county, respectively. Depending on how one defines the urban contours of Beijing, using administrative data alone produces more than one estimate of population within urban Beijing. The dots in the figure represent settlement points, i.e., aspatial representations of urban areas. The United Nations' World Urbanization Prospects and associated database identifies two settlements (green dots inside red dots at Beijing Shi and Langfang) in the Beijing region. GRUMP identifies 11 settlements (red dots) in the same area. The overlay of the NL9495 extent with administrative and point data help to reveal locations missing possible point representation; those places (purple dots) were identified ex post facto after GRUMP. For many GRUMP settlements, it is possible to generate a series of possible comparative population estimates, as noted in the call out boxes. In GRUMP, NL9495 data give an urban space to the point that the UN identifies as Beijing and further combine space between administrative areas. In Beijing county, the NL data identify urban areas that would otherwise be undefined spatially within the county.

floating population is hard to capture through the official census, it is important to consider alternative means for identifying areas of urban change.

Remote sensing can help generate relevant information for policy analysis and research (Liu et al., 2002; Angel et al., 2007). By coupling population data with remote sensing data, it is possible to gain another view of the urban form. The combination of remote sensing and demographic data is a powerful tool for detecting changes that may occur between censuses. In countries with restrictive data-use policies on finely resolved census-type

data, remote sensing data may also provide a means to fill in where a census cannot; such is the case of Beijing. Uninhibited by administrative or political constraints, remote sensing data provide useful urban information where other data sources are restricted.

Remote sensing approaches

Satellite remote sensors, useful for observations of urban environments, operate at frequencies across the electromagnetic spectrum from visible to microwave.

In particular, this entry presents results derived from Optical Linescan System, Landsat, and QuikSCAT satellite data. These datasets have been successfully applied to urban studies.

Night-light

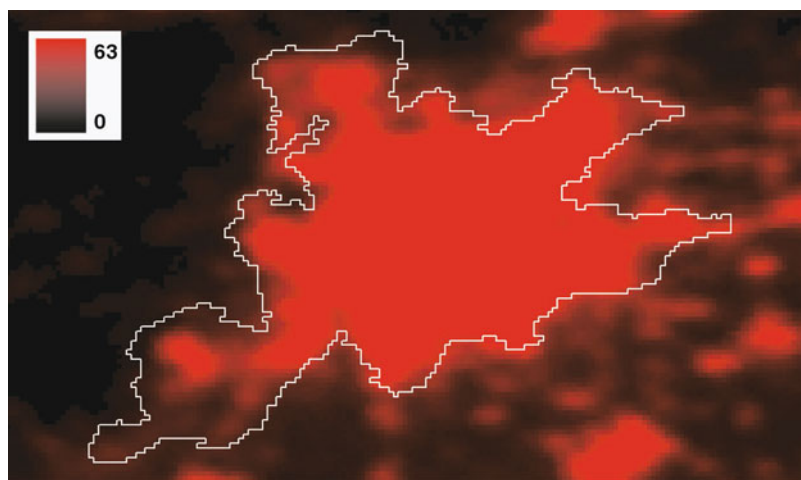
Global NL data, acquired by the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS), have been used as a proxy indicator of urban areas (Elvidge et al., 1997, 2001, 2004; Sutton et al., 1997, 2001; Owen et al., 1998; Balk et al., 2006). The DMSP-OLS was designed to collect global cloud imagery using a pair of broad spectral bands placed in the visible and thermal. The DMSP satellites are flown in polar orbits and each collects 14 orbits per day. With a 3,000 km swath width, each OLS is capable of collecting a complete set of images of the Earth twice a day. At night, the visible band signal is intensified with a photomultiplier tube (PMT) to enable the detection of moonlit clouds. The boost in gain enables the detection of lights present at the Earth's surface. Most of the lights are from human settlements (cities and towns) and ephemeral fires. Gas flares are also detected and can easily be identified when they are offshore or in isolated areas not impacted by urban lighting.

The National Oceanic and Atmospheric Administration (NOAA) National Geophysical Data Center (NGDC) serves as the long-term archive for DMSP measurements, with data extending from 1992 to the present. The archive is organized as individual orbits which are labeled to indicate the year, month, date, and start time. For this study, individual orbits were processed with automatic algorithms that identify image features (such as lights and clouds) and the quality of the nighttime data. The following criteria were used to identify the best nighttime lights data for compositing:

1. Center half of orbital swath (best geolocation, reduced noise, and sharpest features)
2. No sunlight present
3. No moonlight present
4. No solar glare contamination
5. Cloud free (based on thermal detection of clouds)
6. No contamination from auroral emissions

Nighttime image data from individual orbits that meet the above criteria are added into a global latitude–longitude grid (Plate Carrée projection) having a resolution of 30 arc sec. This grid cell size is approximately a square kilometer at the equator. The average visible band digital number (DN) is calculated by dividing the sum of DN values by the number of valid observations. Lights from fires are removed based on their high DN values and low frequency of occurrence. A background mask covering areas devoid of lights is used to define the background noise level, which defines a locally variable threshold for zeroing the digital values in grid cells that do not contain light detections. By overlaying data from two or more years and applying a contrast stretch, it is possible to see changes in lighting over time.

The 2009 NL result for Beijing is presented in Figure 2. NL measurements are acquired globally over several decades with a series of DMSP-OLS. In this study, intercalibrated stable NL data in version 4 are used. For reference, the NL extent contour derived from 1994 to 1995 NL data (called NL9495 hereafter) is overlaid on the 2009 NL image in Figure 2. Including low NL values to sufficiently capture small populated communities, NL9495 is used as a proxy indicator of urban and suburban extent in the Global Rural Urban Mapping Project database (CIESIN et al., 2004; Balk, 2009). Although NL9495 is known to be more expansive than the true extent of a given urban or suburban area due to the



Urban Environments, Beijing Case Study, Figure 2 Intercalibrated stable NL of Beijing from F16 DMSP-OLS satellite measurements in 2009. The color scale is for NL values from 0 (no light) to 63 (maximum saturated intensity).

blooming effects (Elvidge et al., 2004; Small et al., 2005), it is useful as the first demarcation within which a city is amply enclosed. The Beijing change is evident in that the strong NL has expanded even beyond the NL9495 contour limit, especially in the south and in the east of Beijing. While sprawling settlement in Beijing is obvious, the NL in the Beijing center is saturated and consequently the central change is not observable. This limitation can be addressed by multispectral low-light imaging capability of future advanced satellite sensors (Elvidge et al., 2007).

Spectral mixing

Landsat provides the basis for spectral mapping of urban and suburban environments, which are characterized by diversity. A diversity of land uses results in a diversity of land cover types and thus a diversity of spectral characteristics. Analysis of laboratory and field spectra of building materials and impervious surfaces confirms both unique spectral absorptions associated with specific, generally synthetic, materials (Herold et al., 2004; Small, 2009) and the presence of numerous impervious materials that are spectrally indistinguishable from naturally occurring substrates (Small, 2009). In addition to substrates and impervious surfaces, there is also a fundamental ambiguity in the spectral similarity between indigenous and urban vegetation types.

The spectral nonuniqueness of much urban land cover makes it very difficult to accurately delimit urban areas at pixel scales with broadband sensors. Spectral mixing at subpixel scales further complicates the problem. Spatial analysis of high-resolution (1 m) panchromatic imagery from a diversity of urban areas worldwide indicates that the characteristic scale of most urban land cover is in the range of 10–20 m (Small, 2003, 2009). Consequently, almost all pixels imaged by decameter resolution sensors like Landsat are spectrally mixed pixels. Comparative spectral mixture analysis of Landsat 7 multispectral imagery for 30 cities worldwide indicates that there is no unique spectral signature (or range of signatures) that consistently distinguish urban land cover from a wide variety of nonurban land cover types at pixel scales. However, the spectral heterogeneity of urban land cover can often be used to distinguish it from many nonanthropogenic land cover types that are more spectrally homogeneous at pixel scales.

The accuracy of urban land cover classification depends as much on the surrounding nonurban land cover as it does on the urban area itself. In many temperate and tropical environments, urban areas are surrounded by densely vegetated land cover – either indigenous or agriculture. In these cases the urban land cover can often be distinguished by the comparatively low-vegetation densities as many urban environments are characterized by little or no vegetation (Small, 2005, 2009). In arid and semiarid environments, where undeveloped land covers tend to have a variety of exposed rock and soil substrates similar or identical to local building materials,

this distinction is absent, and it becomes much more difficult to accurately map urban land cover with decameter scale multispectral imagery (Small, 2005).

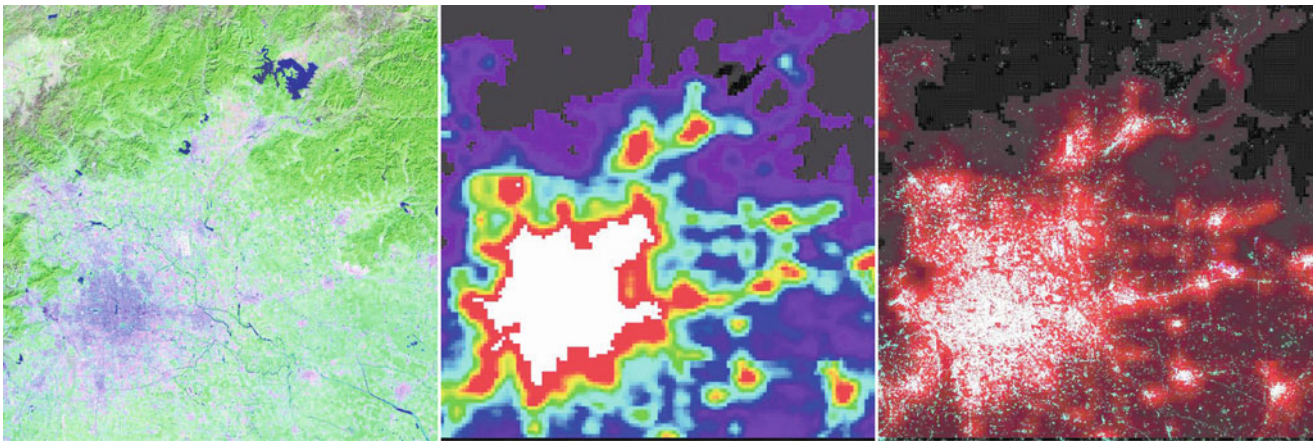
Despite the pervasive spectral ambiguities, it is often possible to derive bounding estimates of urban extent from multispectral imagery. Because many urban areas are characterized by some consistency in building materials, and the pervasive presence of deep shadow, spectral mixtures of substrate and shadow often distinguish urban environments from vegetated surroundings. Starting with training spectra from unambiguously urban core areas, it is possible to incrementally increase the spectral heterogeneity along a mixing trajectory between shadow and substrate to derive bounding estimates of urban extent. Urban extent can often be bounded by deriving conservative and liberal estimates of urban land cover and comparison to higher resolution imagery. The difference in location and spatial extent of these estimates is generally small compared to the average size (Small et al., 2005). Figure 3 shows an example of such bounding estimates for Beijing and how they compare to night-light intensity in 2008. The utility of bounding maps like this is in their ability to characterize the location, size, and approximate extent of urban land cover relative to surrounding land covers which might be classified with greater accuracy. The limitation is the potential for considerable spectral ambiguity and pixel-scale classification errors of both commission and omission.

Scatterometer

Ku-band backscatter data were obtained globally by the SeaWinds scatterometer on the QuikSCAT (QSCAT) satellite since July 1999 (Jet Propulsion Laboratory, 2006). The scatterometer footprint (two-way, half-power, full beamwidth) has an elongated or “egg” shape of about 25 km in azimuth by 37 km in range. Such coarse footprint renders the standard QSCAT data inapplicable to observations of urban areas that require a much higher resolution.

Earlier approaches to increase resolution involve matrix inversion, which is in essence the traditional deconvolution method (DCM). Álvarez-Pérez et al. (2000) assume azimuthally independent radar echo so that DCM can be used; however, the alignment of buildings and man-made structures in cities results in azimuthally dependent backscatter, which violate the validity condition for DCM. DCM further requires that backscatter of the targeted area remains unchanged during the time period of data acquisition for the resolution enhancement (Early and Long, 2001). This additional requirement further invalidates the DCM whenever backscatter changes (e.g., by snowmelt, precipitation, flood, diurnal effects, temperature change, etc.).

To account for the DCM shortfall, DSM (Nghiem et al., 2009) is developed to derive backscatter at a resolution much higher than the 25 km by 37 km footprint. In DSM, the Rosette Transform is applied to both the mean and the



Urban Environments, Beijing Case Study, Figure 3 Urban–rural land cover gradients in reflected and emitted light. Landsat 5 false-color composite (left) in 2008 shows developed, agricultural, and undeveloped land cover. Homogeneous green areas are agriculture while the dark pink, purple, and gray areas are generally associated with mixtures of SWIR-bright building materials and deep-shadow characteristic of more intensive development. Stable night-light brightness (center) for 2008 ranges from dim-low DN values (purple to cyan) to brighter (green to red) and saturated (white). Unlighted areas (black) are undeveloped. Day-night composites (right) combine Landsat and night-lights to illustrate consistencies in land cover and night-light brightness. Superimposed night-light brightness (red) on upper (green) and lower (blue) bounding estimates of built-up extent shows high-density built environments (white) associated with bright night-lights while agriculture and low-density rural land cover are associated with lower light levels (DN < ~20).

fluctuating part of backscatter data collocated at each location within each 1 km² area. The major advantage is that DSM allows the target to change in azimuth and in time to achieve the high resolution using a composition of data over a long period to minimize speckle in the resultant image (the longer the time period the better). All 10 years of QSCAT data can be used together to obtain DSM σ_0 and IV for decadal observations of urban and suburban environment. Moreover, QSCAT data can be partitioned seasonal, annually, or interannually to monitor urban change at different timescales.

DSM is applied to investigate Beijing urban characteristics in this study. The DSM σ_0 (Figure 4), derived from QSCAT data acquired in 2009, shows the extensive Beijing sprawl especially in the southeast side well beyond the NL9495 limit. Compared to the NL 2009 (Figure 2), DSM σ_0 detects more sprawling in the southeast extending into Daxing Xian and Tongzhou. Both DSM σ_0 and NL agree that there is less expansion in the north and the west of Beijing where complex mountains hinder the urban growth.

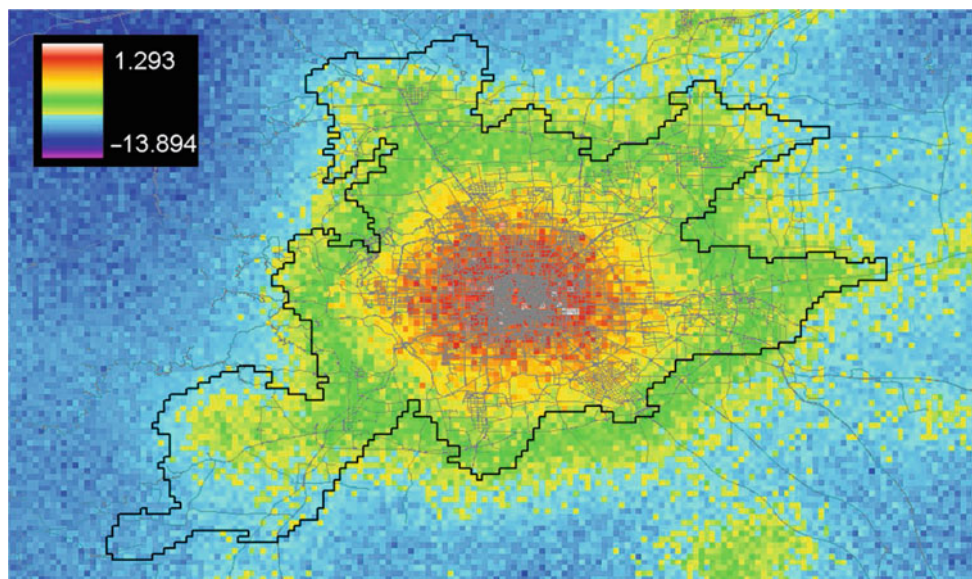
There are small isolated areas identifiable by brighter NL compared to the surroundings (e.g., reddish areas in Daxing Xian and Tongzhou outside the NL9495 limit in the NL image in Figure 2); however, such areas are not distinguished in the DSM σ_0 image (Figure 4). This result suggests that buildings, houses, or other man-made infrastructures, which should have high backscatter, are not substantial in those small communities. This is a limitation of the QSCAT DSM data, and future satellite scatterometers should have a smaller footprint or data should be processed into thinner slices for a higher

resolution so that small human settlements can be identified.

More than just the city extent, the DSM σ_0 image in Figure 4 reveals the core of Beijing (orange to red) with a transition (yellow) into the greater Beijing (green). This urban typology is observable by DSM σ_0 because the mean backscatter has a large dynamic range that can account for radar signatures of different urban areas in Beijing without the saturation problem. The Beijing core identified by DSM σ_0 corresponds to the area of high density of roads (Figure 4). Multiple rings of roads in the Beijing road network encircle the city center with the larger rings closer to the outer part of the Beijing core (orange to red area in Figure 4), which characterize the distinct concentric sprawl, a serious violation of the General Plan of Beijing (Shenghe et al., 2002).

Combining remote sensing and demography

Each satellite sensor has certain capabilities and limitations. To achieve an optimal set of information characterizing urban environments, a combination of multisatellite data together with census data is necessary. The integration of remote sensing data from different satellite sensors in a Geographic Information System (GIS) environment represents a well-established approach being widely recognized as powerful and efficient in monitoring and evaluating urban sprawl and urban land use and land cover change over a specific period of time (Mesev, 1997; Wu et al., 2006; Jat et al., 2008; Feng, 2009). Rindfuss et al. (2003), in particular, highlighted the importance of combining remote sensing data with



Urban Environments, Beijing Case Study, Figure 4 DSM σ_0 of Beijing from QSCAT satellite measurements in 2009. The color scale is for backscatter from low value (magenta) to high value (white). Beijing road network (thin gray lines) is overlaid on the DSM map with road data from CloudMade.com at <http://downloads.cloudmade.com> (accessed on July 1, 2010). The road data are updated weekly (<http://www.openstreetmap.org>).

demographic data, primarily from surveys, and identified it as one of the key aspects in urban land use and land cover change studies (also see Liverman et al., 1998).

DSM-processed QSCAT data provide results evenly gridded at 30 arc sec in latitude and in longitude as a rectangular array of points in a Geographic Coordinate System (GCS) – World Geodetic System (WGS) 84 datum. The reference point is at 0° latitude and 0° longitude, and the rectangular array ranges from 0° to $[360-(1/120)]^\circ$ eastward and from 0° to 88° and -88° northward and southward, respectively. With each point storing its own geographic coordinates, each DSM σ_0 and the associated IV value can be referred to the area (approximately a square kilometer at the equator) it represents. Thus, the rectangular array of points can be used to create a raster grid, with a pixel size of 30 arc sec, in which each cell is identified by the values stored in the point located at its southwest corner (Nghiem et al., 2009).

Each QSCAT raster grid from 2000 to 2009 is coregistered (i.e., spatially aligned so that all pixels representing the same area and location in different rasters are exactly coincidental) with the DMSP-OLS NL raster dataset and with two different population raster datasets modeling the global residential (P_r) and ambient (P_a) population distribution, in order to run a pixel-by-pixel analysis in a GIS environment. Both the Global Rural Urban Mapping Project (GRUMP) dataset representing P_r (CIESIN et al., 2004) and the LandScanTM (LandScan

2008TM/UT-Battelle, LLC) dataset representing P_a (Dobson et al., 2000) depict the distribution of human population across the globe at a resolution of 30 arc sec. These datasets result from methods that transform population data from their native irregularly shaped spatial units (vector format), which are usually administrative and of varying shape, size, and resolution, to a regular raster grid of quadrilateral latitude–longitude cells at a fixed resolution (Tobler et al., 1997; Deichmann et al., 2001; Dobson et al., 2003; Balk and Yetman, 2005). In particular, the allocation mechanism for GRUMP P_r explicitly takes into account population and extent of urban areas (see Balk et al., 2006; Balk, 2009, for a detailed description of the methodology), whereas the allocation mechanism for LS P_a is based on likelihood coefficients calculated considering multiple key indicators of population distribution (see Oak Ridge National Laboratory, 2010, for a description of the general methodology).

In this study, GIS technology is used to process and combine different remote sensing products in order to characterize the Beijing environment in the decade of 2000s (2000–2009) and to highlight the relationship between the remote-sensed “urban sprawl indicators” (NL and QSCAT DSM datasets) and the distribution of the population (as represented both in the GRUMP and in the LS datasets) in the Beijing urban and suburban area. We present the linear correlation results, between the two population raster datasets and both DSM and NL values (Table 1), for Beijing urban and suburban area within the

Urban Environments, Beijing Case Study, Table 1 Correlation analysis between population data (GRUMP 2000 and LS 2004–2008) and remote sensing signatures (DSM backscatter and NL brightness)

	<i>n</i> (5 % cases)	GRUMP 2000		LS 2004–2008	
		DSM 00-09	NL 00-09	DSM 00-09	NL 00-09
Urban area					
Beijing	248	−0.046	−0.12	0.724	0.475

common NL9495 extent. Both the DSM and the NL values used in the analysis were averaged between 2000 and 2009 while the GRUMP and the LS datasets depict, respectively, P_r in the 2000 and P_a averaged over 2004, 2006, 2007, and 2008. Due to the large number of raster grid cells, and because a relatively high number of observations would render even very weak relationships as significant, we use a 5 % random sample of those cells to obtain an appropriate representation of the relationships.

Results from the correlation analysis in Table 1 show a very small correlation coefficient between either DSM σ_0 or NL and P_r from GRUMP 2000. Since infrastructures represented by DSM σ_0 and anthropogenic lights represented by NL must be both physically related to human settlements, the weak negative correlation suggests that P_r is not colocated in the dense built-up areas. In contrast, a strong association is found between DSM σ_0 and P_a from LandScan™ (LS) because houses, offices, factories, or other infrastructures are where people are most likely to be located physically at any time. Similarly, there is an association between NL and P_a . The strong σ_0 – P_a relationship supports the use of radar data in demography and other social sciences.

Figure 5 illustrates the composite approach combining both remote sensing and demographic data to observe urban characteristics of Beijing. The DSM σ_0 , averaged over all data collected in 2000–2009, represents the urban extent and typology in the decade of 2000s. In Figure 5, the spatial pattern of DSM time-average backscatter and that of the LS P_a are similar and thus explain the high correlation between the scatterometer remote sensing data and the ambient population census, whereas the residential population P_r , which is partitioned into each administrative unit, has a spatial distribution different from the scatterometer observation. Larger values (green-yellow, top right panel in Figure 5) of the DSM IV indicate that large changes have occurred in the central core and also in the transition areas of Beijing. The NL image, from data averaged over 2000–2009, has less spatial distinction compared to that of the DSM backscatter within the NL9495; however, NL can detect many more small

settlements outside the NL9495 contour. In contrast to both DSM and NL results, Landsat data reveal a much higher resolution containing more spatial details.

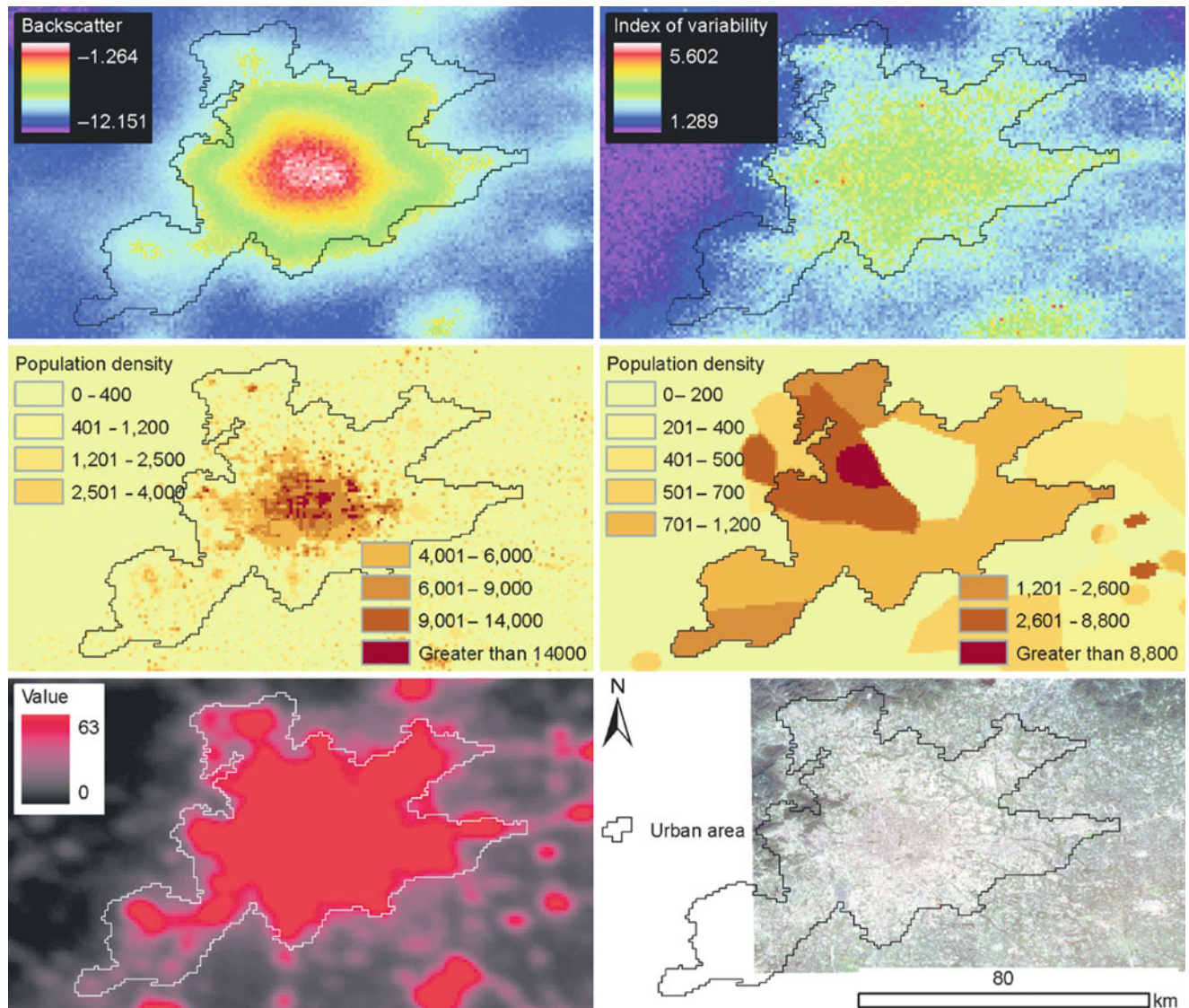
Summary

Various remote sensing methods and demographic datasets are used in the Beijing case study to illustrate their capability to observed physical and demographic characteristics of the urban environment. NL data serve well to identify the outer limit of not only large urban areas but also small settlements. For each large urban contour limit from NL, DSM scatterometer data can detect urban extent and typology. Within each urban type classified by DSM data, Landsat spectral signatures can provide high-resolution details of the urban land cover. It is found that DSM σ_0 has the highest correlation with ambient population of Beijing. To monitor urban change, data can be partitioned into different timescales. The combination of multiple remote sensing methods together with demographic measures is necessary to effectively observe urban environments, rather than each dataset standing alone – both by adding shape and contour to urban population estimates as well as to describe patterns of association between population models and those detecting the rapidly changing built environment. Although Beijing may have local characteristics in detail, it shares many issues of a megacity common to other megacities across the world, where methods and results in this Beijing study can be applicable.

While several important remote sensing approaches applicable to global urban observations are presented here, this Beijing case study is by no means exhaustive. Global synthetic aperture radar (SAR) data such as from the Space Shuttle Radar Mission (STRM) has a great potential to obtain a uniform database of global infrastructure representing the state of global urban environments in the specific period of data collection (Nghiem et al., 2001). With the interferometric SAR capability, the recent TanDEM-X Mission (Krieger et al., 2010) and the future DESDynI Mission (National Research Council, 2007) can provide three-dimensional data to measure building height, while a future class of optical sensors such as that proposed for the Nightsat Mission (Elvidge et al., 2007) can significantly improve the urban remote sensing capability. These advances are critical for developing a consistent global urban infrastructure dataset to meet the need of the social and physical scientific communities as well as policy and decision makers.

Acknowledgments

The research carried out at the Jet Propulsion Laboratory (JPL), California Institute of Technology, was supported under a contract with the National Aeronautics and Space Administration (NASA) Land-Cover and Land-Use Change (LCLUC) Program. The DSM development was



Urban Environments, Beijing Case Study, Figure 5 Composite results for Beijing: *Top left panel* for 2000–2009 DSM average backscatter, *top right panel* for 2000–2009 DSM IV, *middle left panel* for average ambient population P_a from LS 2004–2008, *middle right panel* for residential P_r from GRUMP 2000, *bottom left panel* for 2000–2009 NL, and *bottom right panel* for Landsat. NL9495 contour is overlaid on all images as the common reference area.

funded through the Director's Research and Development Fund (DRDF) program at JPL.

Bibliography

- Álvarez-Pérez, J. L., Marshall, S. J., and Gregson, K., 2000. Resolution improvement of ERS scatterometer data over land by Wiener filtering. *Remote Sensing of Environment*, **71**, 261–271.
- Angel, S., Parent, J., and Civco, D., 2007. Urban sprawl metrics: an analysis of global urban expansion using GIS. In *ASPRS Annual Conference*, Tampa, FL.
- Balk, D., Pozzi, F., Yetman, G., Deichmann, U., and Nelson, A., 2005. The distribution of people and the dimension of place: methodologies to improve the global estimation of urban extents.

In *Proceedings of the Urban Remote Sensing Conference*. Tempe, AZ: International Society for Photogrammetry and Remote Sensing.

- Balk, D., and Yetman, G., 2005. The global distribution of population: evaluating the gains in resolution refinement. Available from World Wide Web: http://beta.sedac.ciesin.columbia.edu/gpw/docs/gpw3_documentation_final.pdf (accessed June 22, 2010).
- Balk, D., Deichmann, U., Yetman, G., Pozzi, F., Hay, S. I., and Nelson, A., 2006. Determining global population distribution: methods, applications and data. In Hay, S. I., Graham, A. J., and Rogers, D. J. (eds.), *Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications*. London: Academic/Advances in Parasitology, Vol. 62, pp. 119–156.

- Balk, D., 2009. More than a name: why is global urban population mapping a GRUMPy proposition? In Gamba, P., and Herold, M. (eds.), *Global Mapping of Human Settlement: Experiences, Data Sets, and Prospects*. New York: Taylor and Francis, pp. 145–161.
- Chengrong, D., Ge, Y., Fei, Z., and Xuehe, L., 2008. Nine trends of changes of China's floating population since the adoption of the reform and opening-up policy. *China Population Today*, **25**(4).
- CIESIN (Center for International Earth Science Information Network) of Columbia University. International Food Policy Research Institute (IPFRI), the World Bank, and Centro Internacional de Agricultura Tropical (CIAT), 2004. *Global Rural–Urban Mapping Project (GRUMP). Urban Extents. Alpha Version*. Palisades, NY: CIESIN, Columbia University. <http://sedac.ciesin.columbia.edu/gpw/documentation.jsp> (accessed June 22, 2010).
- Deichmann, U., Balk, D., and Yetman, G., 2001. Transforming population data for interdisciplinary usages: from census to grid. Available from World Wide Web: <http://sedac.ciesin.columbia.edu/gpw-v2/GPWdocumentation.pdf> (accessed June 22, 2010).
- Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., and Worley, B. A., 2000. A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, **66**(7), 849–857.
- Dobson, J. E., Bright, E. A., Coleman, P. R., and Bhaduri, B. L., 2003. LandScan: a global population database for estimating populations at risk. In Mesev, V. (ed.), *Remotely Sensed Cities*. London: Taylor & Francis, pp. 267–281.
- Early, D. S., and Long, D. G., 2001. Image reconstruction and enhanced resolution imaging from irregular samples. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 291–302.
- Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., and Davis, E. R., 1997. Mapping city lights with nighttime data from the DMSP operational linescan system. *Photogrammetric Engineering and Remote Sensing*, **63**, 727–734.
- Elvidge, C. D., Imhoff, M. L., Baugh, K. E., Hobson, V. R., Nelson, I., Safran, J., Deitz, J. B., and Tuttle, B. T., 2001. Night-time lights of the world: 1994–1995. *ISPRS Journal of Photogrammetry and Remote Sensing*, **56**(2), 81–99.
- Elvidge, C. D., Safran, J., Nelson, I. L., Tuttle, B. T., Hobson, V. R., Baugh, K. E., et al., 2004. Area and position accuracy of DMSP nighttime lights data. In Lunetta, R. S., and Lyon, J. G. (eds.), *Remote Sensing and GIS Accuracy Assessment*. Boca Raton, FL: CRC Press, pp. 281–292.
- Elvidge, C. D., Cinzano, P., Pettit, D. R., Arvesen, J., Sutton, P., Small, C., Nemanis, R., Longcore, T., Rich, C., Safran, J., Weeks, J., and Ebener, S., 2007. The Nightsat mission concept. *International Journal of Remote Sensing*, **28**(12), 2645–2670.
- Feng, L., 2009. Applying remote sensing and GIS on monitoring and measuring urban sprawl. A case study of China. *Revista Internacional Sostenibilidad, Tecnología y Humanismo*, **4**, 47–56.
- Herold, M., Roberts, D. A., Gardner, M. E., and Dennison, P. E., 2004. Spectrometry for urban area remote sensing – development and analysis of a spectral library from 350 to 2400 nm. *Remote Sensing of Environment*, **91**, 304–319.
- Jat, M. K., Garg, P. K., and Khare, D., 2008. Monitoring and modelling of urban sprawl using remote sensing and GIS techniques. *International Journal of Applied Earth Observation and Geoinformation*, **10**(1), 26–43, doi:10.1016/j.jag.2007.04.002.
- Jet Propulsion Laboratory, 2006. *QuikSCAT Science Data Product User's Manual*. Jet Propulsion Laboratory Document D-18053-RevA. Pasadena, CA: JPL.
- Krieger, G., Hajnsek, I., Papathanassiou, K. P., Younis, M., and Moreira, A., 2010. Interferometric synthetic aperture radar (SAR) missions employing formation flying. *Proceedings of the IEEE*, **98**(5), 816–843.
- Liu, S., Prieler, S., and Li, X., 2002. Spatial patterns of urban land use growth in Beijing. *Journal of Geographical Sciences*, **12**(3), 266–274.
- Liverman, D., Moran, E., Rindfuss, R., and Stern, P., 1998. *People and Pixels: Linking Remote Sensing and Social Science*. Washington, DC: National Academy Press.
- Mesev, V., 1997. Remote sensing of urban systems: hierarchical integration with GIS. *Computers, Environment and Urban Systems*, **21**(3/4), 175–187.
- National Research Council, 2007. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. Washington, D.C.: The National Academies Press.
- Nghiem, S. V., Balk, D., Small, C., Deichmann, U., Wannebo, A., Blom, R., Sutton, P., Yetman, G., Chen, R., Rodriguez, E., Houshmand, B., and Neumann, G., 2001. *Global Infrastructure: The Potential of SRTM Data to Break New Ground*. A white paper of the Global Infrastructure Mapping Workshop, JPL D-23049, Pasadena, CA: Jet Propulsion Laboratory, California Institute of Technology, 22 pp.
- Nghiem, S. V., Balk, D., Rodriguez, E., Neumann, G., Sorichetta, A., Small, C., and Elvidge, C. D., 2009. Observations of urban and suburban environments with global satellite scatterometer data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **64**, 367–380, doi:10.1016/j.isprsjprs.2009.01.004.
- Oak Ridge National Laboratory, 2010. *LandScanTM Global Population Database*. Oak Ridge National Laboratory, Geographic Information Science and Technology (GIST) Group, Oak Ridge, Tennessee. Available from World Wide Web: http://www.ornl.gov/sci/landscan/landscan_documentation.shtml (accessed July 29, 2010).
- Owen, T. W., Gallo, K. P., Elvidge, C. D., and Baugh, K. E., 1998. Using DMSP-OLS light frequency data to categorize urban environments associated with US climate observing stations. *International Journal of Remote Sensing*, **19**, 3451–3456.
- Rindfuss, R., Walsh, S., Mishra, V., Fox, J., and Dolcemacolo, G., 2003. Linking household and remotely sensed data: methodological and practical problems. In Fox, J., Rindfuss, R. R., Walsh, S. J., and Mishra, V. (eds.), *People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing*. Boston, MA: Kluwer, pp. 1–29.
- Shenghe, L., Prieler, S., and Xiubin, L., 2002. Spatial patterns of urban land use growth in Beijing. *Journal of Geographical Sciences*, **12**(3), 266–274.
- Small, C., 2003. High resolution spectral mixture analysis of urban reflectance. *Remote Sensing of Environment*, **88**, 170–186.
- Small, C., 2005. Global analysis of urban reflectance. *International Journal of Remote Sensing*, **26**(4), 661–681.
- Small, C., Pozzi, F., and Elvidge, C. D., 2005. Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sensing of Environment*, **96**(3–4), 277–291.
- Small, C., 2009. The color of cities: an overview of urban spectral diversity (Invited chapter). In Herold, M., and Gamba, P. (eds.), *Global Mapping of Human Settlements*. London: Taylor & Francis.
- Sutton, P., Roberts, C., Elvidge, C., and Meij, H., 1997. A comparison of nighttime satellite imagery and population density for the continental United States. *Photogrammetric Engineering and Remote Sensing*, **63**, 1303–1313.
- Sutton, P., Roberts, D., Elvidge, C., and Baugh, K., 2001. Census from Heaven: an estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing*, **22**, 3061–3076.
- Tobler, W., Deichmann, U., Gottsegen, J., and Maloy, K., 1997. World population in a grid of spherical quadrilaterals. *International Journal of Population Geography*, **3**(3), 203–225.
- Wu, Q., Li, H., Wang, R., Paulussen, J., He, Y., Wang, M., Wang, B., and Wang, Z., 2006. Monitoring and predicting land use

change in Beijing using remote sensing and GIS. *Landscape and Urban Planning*, **78**(4), 322–333, doi:10.1016/j.landurbplan.2005.10.002.

Yeh, A. G. O., and Xu, J., 2009. *China's post-reform urbanization: trends and policies*. International Institute for Environment and Development (IIED) – UNFPA Workshop on Population and Urbanization Issues, London, UK.

Cross-references

Data Archival and Distribution
Microwave Radiometers
Radars

URBAN HEAT ISLAND

Lela Prashad
School of Earth and Space Exploration, 100 Cities Project,
Arizona State University, Tempe, AZ, USA

Synonyms

Land use planning; Urban land use; Urban materials

Definition

Urban heat island (UHI): Urban materials such as asphalt, concrete, and bricks predominate in the urban landscape and are most associated with residential, commercial, and industrial land use types. These materials can increase the temperatures of an urban area in comparison to the surrounding rural land, generating an urban heat island (Oke, 2011). Urban materials contribute heat to their surroundings through the way they store and release heat energy from the sun. This effect is especially noticeable at night when streets and buildings continue to radiate heat, elevating minimum temperatures (Goward, 1981) (Figure 1).

Introduction

The typical city on a summer afternoon has been found to be 2.5 °C hotter than its surroundings (Akbari et al., 1990). Different urban materials grouped together by land use type, such as commercial or residential, create patterns of higher and lower temperatures across a city.

The urban heat island (UHI) effect can cause chronic heat stress on people in the urban environment and can lead to higher morbidity and mortality, particularly among vulnerable populations such as the poor and the elderly (Conti et al., 2005; Harlan et al., 2008). The size of an urban area directly impacts the scale of its UHI effect (Oke, 1973).

Although, heat island temperatures generally peak 3–5 h after sunset, elevated urban temperatures also occur in the daytime (Oke, 1982). The most severe heat islands occur during the summer, in cities with little relief, and where there are generally clear skies with little wind. There is also a positive correlation between city size and

the magnitude of an UHI (Oke, 1973). Overall, urban areas with populations of more than 100,000 can be 8–12 °C warmer than rural areas (Bonan, 2002).

Urban researchers and decision makers can utilize remote sensing technology to determine the impact of current land use on the UHI and use future projections of land use plans to mitigate the UHI effect (Voogt and Oke, 2003).

Thermal inertia and surface temperature

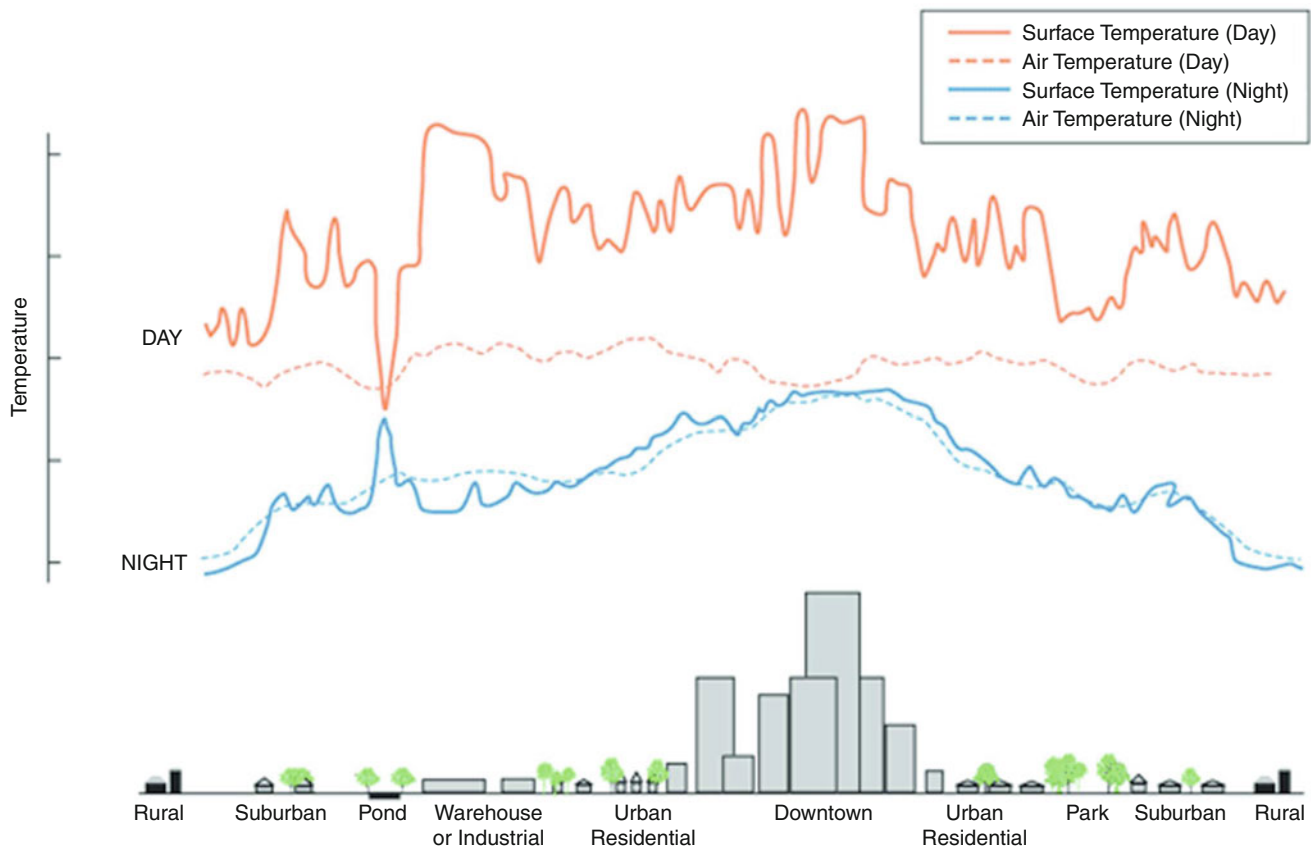
When one thinks of an urban area, one usually pictures tall buildings, streets, highways, living spaces, vegetated yards, and parks laid out in a pattern across the landscape. A city may not seem connected to a natural landscape of rocks and soils but at the molecular level, where thermal properties are defined, there is a strong relationship. Urban materials have similar thermal properties to natural materials and can be studied in the same way through remote sensing.

All materials on the surface of the Earth emit absorbed solar energy, and this emissivity can be recorded by a remote sensor and converted to surface temperature. An urban material's surface temperature is dependent on its physical properties of thermal inertia, albedo (reflectivity), and roughness (Casanova et al., 2007). Urban materials with high emissivities, such as concrete, can absorb high amounts of solar energy. The opposite is true of materials with low emissivities, such as aluminum. Thermal inertia is essentially the ability of a material to conduct and store heat. This means that materials with high thermal inertia such as metal or sand, heat up and cool off quickly, whereas materials with low thermal inertias, such as granite or concrete, heat up slowly and then release their heat slowly, overnight. Roughness, or the three-dimensional structure of an urban area, also plays a role in temperature as tall, densely set buildings create urban canyons that increase the number of surfaces available to absorb heat and block cooling wind currents (Oke, 1987).

Albedo also plays a role in the UHI effect, particularly in relation to the temperature of roofs. Materials with high albedos reflect high levels of solar radiation, while low albedo materials absorb high levels of incoming solar radiation, therefore emitting high amounts of thermal energy. The difference between the surrounding air and low albedo roof coverings can be as high as 50 °C as opposed to 10 °C for high albedo roof coverings (Berdahl and Bretz, 1997). The thermal inertia and albedo of urban materials, in addition to the city structure, work together to generate the UHI effect (Hoyano, 1984; Oke, 1987).

Remote sensing of urban materials

In order to accurately determine emissivity and temperature, a remote sensor must have sufficient spectral resolution in the thermal wavelengths. Instruments such as ASTER, Landsat ETM+, and AVHRR have this capability (Stefanov et al., 2007). ASTER, for example, has five bands in the 8.12–11.65 µm range and can be used to



Urban Heat Island, Figure 1 Temperatures fluctuate over different urban land use types, peaking over the urban core and dipping over water and vegetated areas. This figure shows that while surface temperatures fluctuate widely during the day, they are more constant and match the air temperatures more closely overnight. The urban heat island effect is strongest at night, when minimum nighttime temperatures are elevated compared to those in the surrounding rural area (Voogt, 2000; EPA, 2008).

determine the composition of urban and geologic materials. In addition to emitted energy, reflectance is also necessary to characterize the albedo, vegetation density, and many land cover/land use parameters, such as impervious surfaces.

Choosing the appropriate sensor for a study of urban materials requires an analysis of the spatial, spectral, and temporal resolutions available, as well as cost. For example, airborne instruments often offer very high spatial and spectral resolution (such as the airborne sensor MASTER) but may have low temporal resolution; sometimes only one image is available for a given area. Satellites often have higher temporal resolution (sometimes multiple scenes are available for a single day) and offer global coverage but have lower spatial and spectral resolution than airborne instruments.

Land use and the urban heat Island

Urban areas densely covered by concrete and buildings have higher temperatures than areas with lower building density and abundant vegetation (Hoyano, 1984; Akbari

et al., 1990; Aniello et al., 1995; Voogt, 2000; Dousset and Gourmelon, 2003). Different land use types, such as commercial, residential, or industrial, each contain common urban materials, and these major land use types each have a unique thermal signature (Hoyano, 1984; Voogt, 2000).

Urban materials with high thermal inertias, such as concrete and asphalt, slowly emit heat throughout the night, raising minimum temperatures. For example, in Phoenix, Arizona, USA, summertime lows have increased from 22.7 °C to 26.7 °C over the last 25 years, and overall temperatures are projected to increase by approximately 10° over the next few decades (Balling and Brazel, 1987). This increasing intensity of the UHI can be attributed to population growth and increasing development of land (Cayan and Douglas, 1984; Balling and Brazel, 1986). The downtown and industrial areas of Paris and Los Angeles have seen increased temperatures of greater than 7 °C over the last 10 years. Residential urban areas with high concrete to open land ratios are significantly warmer than areas with less buildings and more vegetation (Hoyano, 1984).

Land use types that have high levels of vegetation, such as parks and open spaces, cool the more densely

developed areas around them through evapotranspiration and can significantly mitigate the UHI effect (Akbari et al., 1990; Dousset and Gourmelon, 2003; Yu and Hien, 2006; Akbari and Shea Rose, 2008; Cao et al., 2010).

Evapotranspiration is a measure of water loss over time from soil and water body evaporation and plant transpiration. Plants and wet soil use a portion of the energy they have absorbed from the sun to convert liquid water into water vapor. This subtracts from the energy that would be radiated from the ground if the ground were bare dirt (Avissar, 1996). The rate of evapotranspiration a plant produces is dependent on how much water the plant requires, the moisture of the soil, and the weather, particularly solar radiation and wind.

Park Cool Island (PCI) effects have been observed to mitigate UHI effects in urban areas. In Sacramento, vegetated parks cooled their surroundings by 5–7 °C through shading and evapotranspiration (Spronken-Smith and Oke, 1998), and parks in Los Angeles and Paris significantly cooled their surroundings (Dousset and Gourmelon, 2003). A study of micro-UHIs in Dallas, Texas, found that areas lacking tree cover were 5–11 °C hotter than those with trees (Aniello et al., 1995).

Heat stress and health

Extreme heat event, exacerbated by the UHI effect, in urban areas has been known to cause chronic heat stress and to affect morbidity and mortality rates (McGeehin and Mirabelli, 2001; Curriero et al., 2002).

Heat stress on a human being can be measured through a human energy budget model which generates a quantitative description of a person's thermal comfort given a certain environment and activity level (Brown and Gillespie, 1995; Heisler and Wang, 2002). People absorb and reflect energy from the sun, just as urban and natural materials do. A person is cooled by sweat evaporated from the skin and wind convection, similar to the evapotranspiration process of a plant.

People of differing socioeconomic groups in urban environments often experience different levels of heat stress due to the different land use and urban materials they are exposed to in their daily lives. Those who are less affluent tend to be exposed to microclimates that have a higher UHI effect (Stefanov et al., 2004; Harlan et al., 2006, 2008). In Phoenix, Arizona, for example, a \$10,000 increase in annual median household income was associated with a 0.28 °C decrease in surface temperature (Jenerette et al., 2007). The elderly are also particularly vulnerable to heat stress exacerbated by the UHI effect. In the summer 2003 heat wave in Italy, for example, there was an increase of 3,134 deaths attributed to heat, 92 % of which were elderly individuals (Conti et al., 2005).

Conclusion

The UHI effect will be a continuing challenge as the world continues to urbanize, cities grow in size and density, and

are increasingly also impacted by global climate change. Satellite remote sensing can play a significant role in monitoring the land use and material properties that contribute to the UHI effect and can provide a common platform to compare UHI effects across cities. Both urban researchers and decision makers currently have access to free and low-cost remote sensing data they can use to quantify the impact of the UHI on their cities, such as Landsat, MODIS, and ASTER. Future global satellite remote sensing instruments higher thermal infrared spectral resolution, with the ability to acquire nighttime data, and finer spatial and temporal resolutions will be needed to further quantify and compare UHI effects globally.

Cities can currently take steps to mitigate their UHI effects through land use planning strategies, such as using high albedo (reflective) roofs and road surfaces, strategically locating parks and shade trees, and using concretes and building materials with lower heat capacities. Cities can now also receive guidance on the public health implications of the UHI effect and produce land use plans that take into account the heat stress on its citizens, particularly vulnerable populations such as the poor and the elderly.

Bibliography

- Akbari, H., and Shea Rose, L., 2008. Urban surfaces and heat island mitigation potentials. *Journal of Human-Environment Systems*, **11**, 85–101.
- Akbari, H., Rosenfeld, A., and Taha, H., 1990. Summer heat islands, urban trees, and white surfaces. *ASHRAE Transactions*, **96**, 1381–1388.
- Aniello, C., Morgan, K., Busbey, A., and Newland, L., 1995. Mapping micro-urban heat islands using Landsat TM and a GIS. *Computers and Geosciences*, **21**, 965–967.
- Avissar, R. 1996. Potential effects of vegetation on the urban thermal environment. *Atmospheric Environment*, **30**(3), 437–448.
- Balling, R., and Brazel, S., 1986. Temporal analysis of summertime weather stress levels in Phoenix, Arizona. *Meteorology and Atmospheric Physics*, **36**, 331–342.
- Balling, R. C., and Brazel, S. W., 1987. Time and space characteristics of the Phoenix urban heat island. *Journal of the Arizona-Nevada Academy of Science*, **21**, 75–81.
- Berdahl, P., and Bretz, S. E., 1997. Preliminary survey of the solar reflectance of cool roofing materials. *Energy and Buildings*, **25**, 149–158.
- Bonan, G., 2002. Urban ecosystems. In *Ecological Climatology*. Cambridge: Cambridge University Press, pp. 547–582.
- Brown, R. D., and Gillespie, T. J., 1995. *Microclimatic Landscape Design: Creating Thermal Comfort and Energy Efficiency*. New York: Wiley.
- Cao, X., Onishi, A., Chen, J., and Imura, H., 2010. Quantifying the cool island intensity of urban parks using ASTER and IKONOS data. *Landscape and Urban Planning*, **96**, 224–231.
- Casanova, G., Bertrand, C., and De Ridder, K., 2007. Measuring urban surfaces' thermal inertia. Paper presented at the *EUMETSAT Meteorological Satellite Conference and the 15th Satellite Meteorology & Oceanography Conference of the American Meteorological Society*, Amsterdam, The Netherlands, September 24–28, 2007.
- Cayan, D. R., and Douglas, A. V., 1984. Urban influences on surface temperatures in the southwestern United States during recent decades. *Journal of Applied Meteorology*, **23**, 1520–1530.

- Conti, S., Meli, P., Minelli, G., Solimini, R., Toccaceli, V., Vichi, M., Beltrano, C., and Perini, L., 2005. Epidemiologic study of mortality during the Summer 2003 heat wave in Italy. *Environmental Research*, **98**, 390–399.
- Curriero, F. C., Heiner, K. S., Samet, J. M., Zeger, S. L., Strug, L., and Patz, J. A., 2002. Temperature and mortality in 11 cities of the eastern United States. *American Journal of Epidemiology*, **155**, 80.
- Dousset, B., and Gourmelon, F., 2003. Satellite multi-sensor data analysis of urban surface temperatures and landcover. *ISPRS Journal of Photogrammetry and Remote Sensing*, **58**, 43–54.
- EPA, U. S., 2008. *Reducing Urban Heat Islands: Compendium of Strategies Urban Heat Island Basics*. Washington, DC: Office of Atmospheric Programs.
- Goward, S. N., 1981. Thermal behavior of urban landscapes and the urban heat island. *Physical Geography*, **2**, 19–33.
- Harlan, S. L., Brazel, A. J., Prashad, L., Stefanov, W. L., and Larsen, L., 2006. Neighborhood microclimates and vulnerability to heat stress. *Social Science & Medicine*, **63**, 2847–2863.
- Harlan, S. L., Brazel, A. J., Jenerette, G. D., Jones, N. S., Larsen, L., Prashad, L., and Stefanov, W. L., 2008. In the shade of affluence: the inequitable distribution of the urban heat island. *Research in social problems and public policy*, **15**, 173–202.
- Heisler, G. M., and Wang, Y., 2002. Applications of a human thermal comfort model. In *Fourth Symposium on the Urban Environment*. Boston, MA: American Meteorological Society.
- Hoyano, A., 1984. Relationships between the type of residential area and the aspects of surface temperature and solar reflectance (based on digital image analysis using airborne multispectral scanner data). *Energy and Buildings*, **7**, 159–173.
- Jenerette, G. D., Harlan, S. L., Brazel, A., Jones, N., Larsen, L., and Stefanov, W. L., 2007. Regional relationships between surface temperature, vegetation, and human settlement in a rapidly urbanizing ecosystem. *Landscape Ecology*, **22**, 353–365.
- McGeehin, M. A., and Mirabelli, M., 2001. The potential impacts of climate variability and change on temperature-related morbidity and mortality in the United States. *Environmental Health Perspectives*, **109**, 185.
- Oke, T. R., 1973. City size and the urban heat island. *Atmospheric Environment*, **7**, 769–779.
- Oke, T., 1982. The energetic basis of the urban heat island. *The Quarterly Journal of the Royal Meteorological Society*, **108**, 1–24.
- Oke, T. R., 1987. *Boundary Layer Climates*, 2nd edn. London/New York: Methuen.
- Oke, T., 2011. Urban heat islands. In *The Routledge Handbook of Urban Ecology*, New York, NY: Routledge p. 120.
- Spronken-Smith, R., and Oke, T., 1998. The thermal regime of urban parks in two cities with different summer climates. *International Journal of Remote Sensing*, **19**, 2085–2104.
- Stefanov, W. L., Prashad, L., Eisinger, C., Brazel, A., and Harlan, S. L., 2004. Investigation of human modifications of landscape and climate in the phoenix Arizona metropolitan area using MASTER data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **35**, 1139–1347.
- Stefanov, W. L., Netzband, M., Moller, M. S., Redman, C. L., and Mack, C., 2007. Phoenix, Arizona, USA: applications of remote sensing in a rapidly urbanizing desert region. In Netzband, M., Stefanov, W. L., and Redman, C. (eds.), *Applied Remote Sensing for Urban Planning, Governance and Sustainability*. Berlin: Springer, pp. 137–164.
- Voogt, J., 2000. Image representations of complete urban surface temperatures. *Geocarto International*, **15**, 21–32.
- Voogt, J., and Oke, T., 2003. Thermal remote sensing of urban climates. *Remote Sensing of Environment*, **86**, 370–384.
- Yu, C., and Hien, W. N., 2006. Thermal benefits of city parks. *Energy and Buildings*, **38**, 105–120.

Cross-references

[Climate Monitoring and Prediction](#)
[Thermal Radiation Sensors \(Emitted\)](#)

V

VEGETATION INDICES

Alfredo Huete
Plant Functional Biology and Climate Change Cluster,
Faculty of Science, University of Technology, Sydney,
NSW, Australia

Synonyms

Greenness

Definition

Greenness. The composite optical property of canopy chlorophyll content, leaf area, vegetation cover, and structure.

Soil line. A scatterplot of soil spectra in red and NIR space that forms a zero vegetation baseline from which vegetation can be more accurately measured.

Introduction

Vegetation indices are optical measures of vegetation canopy “greenness” used to quantify vegetation amounts and vigor. A VI measurement combines the chlorophyll-absorbing *red* spectral region (0.6–0.7 μm) with the nonabsorbing, leaf reflectance signal in the *near-infrared* (NIR; 0.7–1.3 μm) to provide a consistent and robust measure of area-averaged canopy photosynthetic capacity. The first vegetation indices were formed by spectral ratioing of the NIR with red band using field radiometers. The greater the amount or vigor of green vegetation present, the larger the contrast between NIR and red reflectances and the higher the resulting NIR/red ratio values, while a reduced spectral contrast would be attributed to lower vegetation amounts, senescing vegetation, or vegetation under stress.

Satellite-based vegetation indices (VIs) were developed in the 1970s for landscape monitoring and have

become indispensable tools in a wide range of applications and decision support systems for scientific and societal purposes. Example applications include natural resource monitoring, agricultural efficiency and famine early warning systems, carbon and water management, biodiversity and invasive species assessments, and environmental health. VIs were operationally implemented with the normalized difference vegetation index (NDVI) global data record computed from the NOAA Advanced Very High Resolution Radiometers (AVHRR) since 1981 (Brown et al., 2006). Many global scientific advancements, related to biogeochemical cycling, net primary production, and vegetation responses to climate variability, can be attributed to the development and consistency of the AVHRR–NDVI time series record (Cracknell, 1997).

Vegetation index equations

There are many ways in which the red, NIR, and other bands may be combined to measure vegetation, and this has resulted in a multitude of VI equations that include band ratios, normalized differences, linear band combinations, and optimized band combinations (Tucker, 1979). The NDVI is a functional variant of the NIR/red band ratio that provides greenness values between -1 and $+1$:

$$\text{NDVI} = (\rho_{\text{N}}/\rho_{\text{R}} - 1)/(\rho_{\text{N}}/\rho_{\text{R}} + 1) = (\rho_{\text{N}} - \rho_{\text{R}})/(\rho_{\text{N}} + \rho_{\text{R}}) \quad (1)$$

where $\rho_{\text{N,R}}$ are reflectances in the NIR and red bands, respectively. An important advantage of the NDVI, as a ratio, is its ability to produce stable values by normalizing many extraneous sources of noise. This was particularly useful in the early AVHRR era, in which data quality problems due to sensor instrument noise, illumination differences, and atmosphere contamination were particularly prevalent.

There are some disadvantages with NDVI in landscape studies, related to the nonlinear behavior of ratios, sensitivity to soil background, and saturation at moderate to high vegetation densities. As a result, optimized variants of the NDVI have been developed that capture essential biophysical phenomena with improved fidelity and with minimal external influences. For example, the soil-adjusted vegetation index (SAVI) was developed to minimize soil influences by incorporating an adjustment parameter (L) in the NDVI equation:

$$\text{SAVI} = (1 + L)(\rho_N - \rho_R) / (L + \rho_N + \rho_R) \quad (2)$$

Other variants of SAVI include the transformed SAVI (TSAVI), the generalized SAVI (GESAVI), and the modified SAVI (MSAVI) and are reviewed in Jiang et al. (2007).

VIs have also been optimized for resistance to atmosphere influences, as in the atmospherically resistant vegetation index (ARVI) and the enhanced vegetation index (EVI), which utilize a blue band to correct for aerosol effects in the red band (Huete et al., 2002). The global environment monitoring index (GEMI) was designed to be resistant to atmosphere effects in AVHRR–NDVI data sets (Pinty and Verstraete, 1992). The EVI minimizes soil background influences and reduces VI signal saturation problems encountered over densely vegetated forests and croplands. The wide dynamic range vegetation index (WDRVI) also enhances sensitivity over highly vegetated conditions (Gitelson, 2004). Increased sensitivity is achieved through optimized use of the NIR band, which exhibits greater optical penetration in high-biomass areas.

Quantitative measures of greenness may also be formulated through linear combinations of the red and NIR bands, as in the perpendicular vegetation index (PVI) (Richardson and Wiegand, 1977) that measures greenness as the orthogonal distance of a pixel to a “soil line” in red–NIR space. This concept is readily extended to multiple-band space whereby the spectral structure of vegetated pixels resembles a three-dimensional “tasseled cap” shape defined by a soil plane, a greenness axis, and either a yellowness axis (plant senescence) or a *wetness* axis, depending on the bands utilized (Crist and Kauth, 1986).

Biophysical properties and processes

VIs are not *intrinsic* physical quantities themselves, but are widely used as proxies in estimating canopy state variables, including leaf area index (LAI), fraction of absorbed photosynthetically active radiation (f_{APAR}), chlorophyll content, and vegetation fraction (F_v). VI relationships with these biophysical quantities have been extensively studied through field measurements and with canopy radiative transfer models (Huemmrich, 2001) and found to vary greatly over the variable structural assemblages of foliage cover, leaf area, and chlorophyll typically found across different land cover types and phenology stages. As a consequence, satellite studies

show VIs are often only moderately correlated with specific canopy variables, are land cover dependent, and not easily extended to regional and global scales that involve multiple vegetation types and canopy structures. The numerous applications of satellite VIs in ecological studies are extensively reviewed in Glenn et al. (2008), Kerr and Ostrovsky (2003), and Pettorelli et al. (2005).

VIs are found to be strongly related to green vegetation fraction (F_v), although both linear and nonlinear relationships are reported (Carlson and Ripley, 1997). Green vegetation fraction can also be derived through simple linear combinations of high (vegetation) and low (soil) VI values within a scene or biome type. In the case of LAI, there is a general positive association with VIs at low LAI values between 0 and 2 and increasingly weak sensitivity at LAI beyond 2 or 3 (Baret and Guyot, 1991), although increased sensitivity in higher LAI canopies has been reported with the EVI in dense (LAI > 5) Amazon forests (Huete et al., 2006). Canopy chlorophyll content is an important plant community property in studies of ecosystem productivity, CO_2 fluxes, and vegetation stress. VI relationships with total canopy chlorophyll content tend to be curvilinear related because surface leaves intercept more light than leaves deeper in the canopy. However, the direct retrieval of chlorophyll content from satellites is confounded by other factors, such as canopy architecture, chlorophyll distribution within the canopy, LAI, and soil background. Vegetation canopy f_{APAR} measurements are widely used in radiation dynamics and light-use efficiency studies to relate satellite VIs with primary productivity and biomass accumulation. The NDVI has been shown to be linearly correlated with f_{APAR} in field and modeling studies, and its integral over the growing season has been correlated with ecosystem net primary production (Baret and Guyot, 1991). In a multi-biome study across North America, Sims et al. (2006) empirically derived a linear f_{APAR} –NDVI relationship across 16 different species as $f_{\text{APAR}} = 1.24 * \text{NDVI} - 0.168$.

Ecosystem processes

Recent applications of VIs have demonstrated their utility in studies of ecosystem functioning and in deriving estimates of the rates of ecosystem processes, such as photosynthesis and transpiration that depend on PAR light absorbed by leaves (Glenn et al., 2008). Vegetation gross primary productivity (GPP) is related to the amount of canopy absorbed PAR, or APAR, through a light-use efficiency (LUE) term, specific to vegetation type, with which the absorbed light is used in carbon fixation, or photosynthesis. This is depicted as $\text{LUE} = \text{GPP}/\text{APAR}$ (Monteith and Unsworth, 1990).

Satellite VI data have been shown useful in providing repetitive observations of seasonal and interannual variations in vegetation activity in response to climate and land use forcings. VIs are widely used in characterizing various vegetation phenophases, such as the onset of greening,

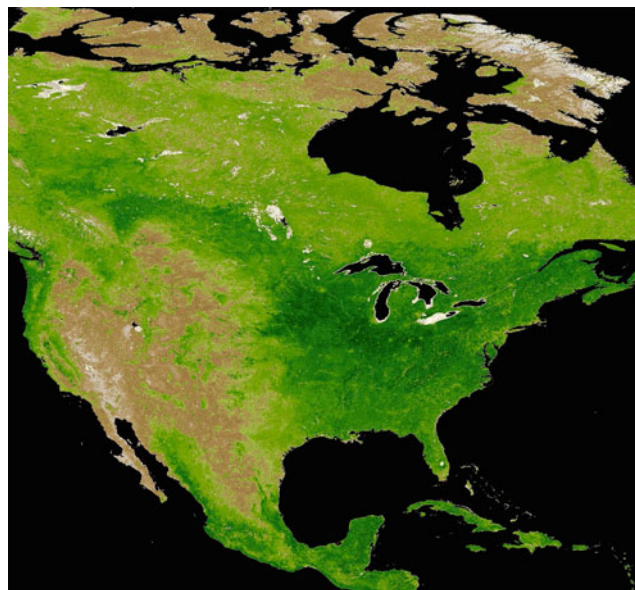
leaf development, length of the growing season, and subsequent browning and dormancy, which are critical to understanding ecosystem functioning and associated seasonal patterns of carbon, water, and energy fluxes (Zhang et al., 2003).

Moderate and coarse resolution vegetation index products

There now exist multiple satellite VI products, generated from the AVHRR instruments and from newer sensor systems, including the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors onboard the Terra and Aqua Earth Observing System platforms, the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), and SPOT-VEGETATION instruments. Vegetation indices are seamless data products that are computed the same across all pixels in time and space, without prior assumptions of biome type, land cover condition, or soil type, and thus represent actual, long-term measurements of the land surface. They are readily fused and intercalibrated across multiple sensor systems to maintain critical long-term time series data needed for climate-related process studies. Satellite VI products work optimally with cloud filtering, radiometric calibration, precise geolocation, and a snow mask. In addition, the VI product performs best using atmosphere-corrected surface reflectance inputs that are further *composited* over set time intervals for removal of cloud and cloud shadow contamination and to constrain sensor view angles and minimize the presence of bad quality pixels (Figure 1). VI data sets are available at varying spatial (250 m to 8 km) and temporal (daily to monthly) resolutions to capture essential vegetation biophysical dynamics and to meet the needs of the user community.

The accuracy of these VI products will vary in space and time as a result of spatial and seasonal variations in cloud persistence, unresolved clouds, and quality of upstream surface reflectance retrievals. Validation and accuracy assessments of satellite VI products are quite challenging and require periodic samplings of vegetation states at scales on the order of a kilometer. Large-scale field campaigns, such as HAPEX, BOREAS, and LBA, provide valuable ground validation data over specific biome types and climates. Regional and global tower networks equipped with NDVI-like hemispherical sensors and eddy covariance flux measures of CO₂ and H₂O offer more validation opportunities over a more complete sampling of land cover types. Phenology networks also provide budbreak, greening, and browning information that will improve performance assessments of VI satellite products.

Direct broadcast and readout technologies also provide near real-time access to satellite imagery and VI products for a variety of rapid response applications, including fire detection and disaster mitigation (<http://directreadout.sci.gsfc.nasa.gov/>). A quick turnaround time between data acquisition and VI product delivery greatly enhances the



Vegetation Indices, Figure 1 Example of the enhanced vegetation index (EVI) computed from 16 day composited (April 7–22, 2003) Terra-MODIS reflectances in bands 1 (red), 2 (NIR), and 3 (blue) for a portion of North America (the darker green colors indicate highest vegetation densities).

benefit for many user groups that need to respond quickly to observed changes in vegetation activity due to drought, fire, and extreme weather events.

Conclusion

As a result of their simplicity, vegetation indices are among the most robust techniques in remote sensing, yielding consistent spatial and seasonal to long-term comparisons of green vegetation at local to global scales. VIs are invaluable tools in monitoring vegetation biophysical conditions related to green foliage density and ecosystem processes related to light absorbed by green vegetation, specifically canopy photosynthesis and plant transpiration. There exist many VIs and all have been related to some extent with various biophysical quantities; however, there remain significant differences in how they depict “greenness,” rendering multiple VIs potentially useful and complementary.

Bibliography

- Baret, F., and Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, **35**, 161–173.
- Brown, M. E., Pinzon, J. E., Didan, K., Morisette, J. T., and Tucker, C. J., 2006. Evaluation of the consistency of long-term NDVI Time series derived from AVHRR, SPOT-Vegetation, SeaWiFS, MODIS, and Landsat ETM + Sensors. *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 1787–1793.

- Carlson, T. N., and Ripley, D. A., 1997. On the relation between NDVI, fractional vegetation cover and leaf area index. *Remote Sensing of Environment*, **62**, 241–252.
- Cracknell, A. P., 1997. *The Advanced Very High Resolution Radiometer (AVHRR)*. London: Taylor & Francis. 534p.
- Crist, E. P., and Kauth, R. J., 1986. The tasseled cap de-mystified. *Photogrammetric Engineering and Remote Sensing*, **52**, 81–86.
- Gitelson, A. A., 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology*, **161**, 165–173.
- Glenn, E. P., Huete, A. R., Nagler, P. L., and Nelson, S. G., 2008. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: what vegetation indices can and cannot tell us about the landscape. *Sensors*, **8**, 2136–2160.
- Huemmerich, K. F., 2001. The GeoSail model: a simple addition to the SAIL model to describe discontinuous canopy reflectance. *Remote Sensing of Environment*, **75**, 423–431.
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, **83**, 195–213.
- Huete, A. R., Didan, K., Shimabukuro, Y. E., Ratana, P., Saleska, S. R., Hutya, L. R., Yang, W., Nemani, R. R., and Myneni, R., 2006. Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters*, doi:10.1029/2005GL025583.
- Jiang, Z., Huete, A. R., Li, J., and Qi, J., 2007. Interpretation of the modified soil-adjusted vegetation index isolines in red-NIR reflectance space. *Journal of Applied Remote Sensing*, **1**, 013503–013512.
- Kerr, J. T., and Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, **18**, 299–305.
- Monteith, J. L., and Unsworth, M. H., 1990. *Principles of Environmental Physics*. London: Arnold. 291p.
- Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., and Stenseth, N. C., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology and Evolution*, **20**, 503–510.
- Pinty, B., and Verstraete, M. M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetation*, **101**, 15–20.
- Richardson, A. J., and Wiegand, C. L., 1977. Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, **43**, 1541–1552.
- Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Flanagan, L. B., Goldstein, A. H., Hollinger, D. Y., Misson, L., Monson, R. K., Oechel, W. C., Schmid, H. P., Wofsy, S. C., and Xu, L., 2006. On the use of MODIS EVI to assess gross primary productivity of North American ecosystems. *Journal of Geophysical Research*, **111**, G04015, doi:04010.01029/02006JG000162.
- Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**, 127–150.
- Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., and Huete, A., 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, **84**, 471–475.

Cross-references

[Land-Atmosphere Interactions](#), [Evapotranspiration](#)
[Vegetation Phenology](#)
[Water and Energy Cycles](#)

VEGETATION PHENOLOGY

John Kimball

Flathead Lake Biological Station, University of Montana,
 Polson, MT, USA

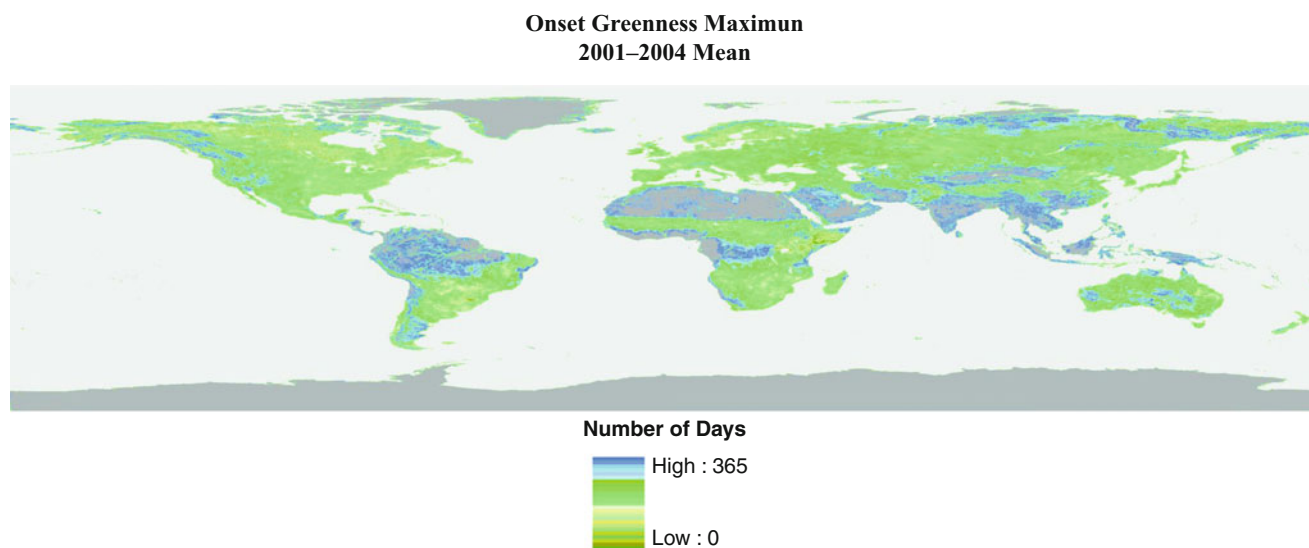
Definition

Vegetation phenology is the timing of seasonal developmental stages in plant life cycles including bud burst, canopy growth, flowering, and senescence, which are closely coupled to seasonally varying weather patterns. Land surface phenology (LSP) is a related term describing vegetation changes observed from satellite remote sensing, recognizing that satellite observations constitute an aggregate signal from heterogeneous surface conditions that may not be representative of any single plant species response. Thus, the satellite LSP measurement is an appropriate indicator of broadscale vegetation-climate interactions and associated landscape changes in ecosystem condition.

Introduction

Vegetation phenology strongly influences land-atmosphere energy, water, and trace gas exchanges. Canopy growth and senescence affect surface albedo and sensible and latent energy partitioning, with direct impacts to weather and climate. Vegetation phenology both influences and is responsive to environmental conditions including temperature, moisture supply, humidity, and photoperiod. Phenology events are spatially complex and temporally dynamic, influenced by vegetation morphology, species, and microclimate variability. Plant phenology in temperate and extratropical latitudes coincides with seasonal changes in day length and temperature, whereas tropical forest, savanna, and grassland phenology varies with seasonal wet/dry events and can display multiple periods of growth and senescence during a single annual cycle.

Satellite remote sensing provides for systematic, repeat global observations that are sensitive to vegetation conditions. Potential uses of land surface phenology (LSP) information include defining vegetation canopy status in large-scale climate and biogeochemical models, and analyses of climate-biosphere interactions. Numerous phenology measures have been derived from satellite remote sensing time series (e.g., Reed et al., 2003), including the following: time of growing season onset (initiation of sustained photosynthesis), offset (canopy senescence, leaf drop, and dormancy), and duration; date of maximum greenness or canopy development; seasonal amplitude of vegetation greenness (magnitude of vegetation canopy growth); and rates of greenup (photosynthetic increase) and senescence (photosynthetic decrease). The most appropriate satellite measurements for global LSP monitoring provide nearly complete coverage of the terrestrial biosphere and frequent (~weekly) sampling to resolve rapid changes in vegetation life cycles. Appropriate spatial



Vegetation Phenology, Figure 1 Mean annual maximum vegetation greenness onset from the MODIS (MOD12Q2) phenology product. *Gray* areas represent frequent cloud cover, open water, ice, and other nonvegetated land masked from further processing; *dark blue* areas represent vegetation with low seasonal EVI range.

scales for LSP assessment approach 30 m resolution for characterizing individual vegetation patches, while coarser scale (>1 km resolution) observations are sufficient to characterize climate-biosphere interactions. Finer scale mapping is constrained by the need for high temporal fidelity and global coverage. Most satellite remote sensing studies utilize optical or microwave regions of the electromagnetic spectrum for LSP assessment, each with associated trade-offs in resolution and sensitivity to vegetation and atmospheric parameters.

Optical remote sensing

Most optical remote sensing LSP studies exploit the sensitivity of red and near-infrared wavelengths to photosynthetic biomass, including presence, abundance, and distribution of green leaves. The potential for LSP monitoring from space was recognized with the initiation of Landsat in the 1970s, though the 16–18 day repeat cycle of these measurements was generally too coarse to resolve rapid vegetation changes. Initiation of the NOAA Advanced Very High Resolution Radiometer (AVHRR) series in the early 1980s made LSP monitoring practical on a global basis. Many subsequent satellite studies of global vegetation dynamics have used the AVHRR record. However, the NASA Moderate Resolution Imaging Spectroradiometer (MODIS) record has rapidly become a sensor of choice for vegetation studies due to improved calibration, advanced cloud detection and aerosol screening capabilities, and heightened spectral sensitivity to photosynthetic biomass. The MODIS record includes an operational (MOD12Q2) LSP product (https://lpdaac.usgs.gov/products/modis_products_table) derived using enhanced vegetation index (EVI) time series from the

EOS Terra satellite; the MOD12Q2 product is produced twice annually at 1 km resolution and includes multiple variables for each grid cell, including timing of maximum vegetation greenness onset (Figure 1). A closely related MYD12Q2 LSP product is derived using MODIS EVI data from both Terra and Aqua satellites and is produced annually at finer (500 m) resolution. These data provide a continuous (from 2000) global record of seasonal to decadal variability in vegetation greenness and associated LSP dynamics suitable for climate change studies.

Green vegetation shows preferential absorption and reflection in respective red and near-infrared portions of the visible spectrum, which forms the basis of the LSP measurement from optical remote sensing. A common vegetation index (VI) for phenology studies is the Normalized Difference Vegetation Index (NDVI), which is the ratio of the difference between red and near-infrared bands over their sum. The NDVI is sensitive to changes in vegetation greenness, while minimizing variable atmosphere, illumination, and soil background effects. Other VIs include the EVI which provides improved vegetation discrimination and reduced signal saturation over dense vegetation; a VI derivative is the Leaf Area Index (LAI) which provides a measure of photosynthetic leaf area that accounts for variable canopy structure.

Optical remote sensing of vegetation has several limitations including atmosphere cloud/aerosol contamination, varying soil/surface background reflectivity, and seasonal reductions in solar illumination at higher latitudes. Efforts to mitigate atmospheric effects include spatial/temporal maximum value compositing (MVC) of the VI time series; commonly used satellite data records employing the MVC approach include AVHRR NDVI Global Inventory Modeling and Mapping Studies (GIMMS) 15 day (<http://glcf>.

umiacs.umd.edu/data/gimms/) composites. The lack of precise calibration and cloud screening promote higher noise levels in the AVHRR record, though the GIMMS NDVI product is derived using extensive post processing of the data to mitigate sensor calibration, view geometry, volcanic aerosols, and other known effects not related to vegetation. The MODIS VI and LAI 8- and 16 day global products are derived using higher-quality daily retrievals and variable (including MVC) compositing techniques depending on the number and quality of observations. MODIS incorporates advanced cloud screening and atmospheric aerosol corrections and provides an improved basis for LSP monitoring. However, the resulting VI time series can be too infrequent to resolve temporally dynamic phenology events. Other limitations include signal degradation from reduced solar illumination and shadowing at high latitudes. The presence of snow can decrease the VI independent of changes in photosynthetic biomass. Missing or degraded satellite observations during critical vegetation phases also reduce LSP accuracy.

Microwave remote sensing

Satellite active (radar) and passive microwave sensors can detect LSP-related changes in vegetation canopy structure and biomass, independent of many environmental constraints on optical remote sensing. Satellite microwave sensors used for LSP studies include the NASA SeaWinds scatterometer and Advanced Microwave Scanning Radiometer for EOS (AMSR-E), which are capable of near daily global monitoring and have relatively long data records. Satellite observations at lower ($\sim \leq 37$ GHz) microwave frequencies are insensitive to solar illumination, clouds, and other atmosphere effects; these attributes enable potential global monitoring day or night and under virtually all weather conditions, dependent upon on sensor design and orbital geometry.

Satellite microwave LSP assessments include temporal change classifications of radar backscatter or brightness temperature (T_b) series to determine landscape freeze/thaw status and associated frozen temperature constraints to canopy gas exchange and vegetation growing seasons (Kimball et al., 2004). SeaWinds Ku-band radar backscatter data was used to estimate canopy LAI and associated LSP changes over a range of global vegetation types (Frolking et al., 2006), while a daily emissivity difference index of 19 and 37 GHz channels from the Special Sensor Microwave/Imager (SSM/I) was used to detect spring onset and canopy development over broadleaf deciduous forest (Min and Lin, 2006). The vegetation optical depth (VOD) retrieval from passive microwave remote sensing is a frequency-dependent measure of canopy extinction of land surface microwave emissions; AMSR-E 10.7 GHz frequency VOD retrievals were used to map the LSP start of season over North America and document relations between MODIS VI greenup and VOD-derived canopy biomass changes for different ecoregions (Jones et al., 2012).

Vegetation absorbs, scatters, and emits microwave energy; the resulting LSP signal is strongly dependent on sensing frequency and polarization, and vegetation canopy structure and biomass (Ulaby et al., 1986). The vegetation canopy signal is significant at Ku-, X-, and C-band frequencies (4–18 GHz) with wavelengths similar to the characteristic (1.7–7.5 cm) size of individual leaves. The signal contribution from deeper canopy layers, including photosynthetic and non-photosynthetic components, is larger at lower microwave frequencies, while sensitivity to land surface conditions is inversely proportional to overlying vegetation biomass density.

A limitation of passive microwave remote sensing for LSP monitoring is a relatively coarse sensor footprint compared to radar and optical sensors. Spatial resolution of microwave radiometers is frequency dependent and varies from approximately 30 km (37 GHz) to 60 km (19 GHz) for the SSM/I (Special Sensor Microwave/Imager). Terrestrial emissions of microwave energy are naturally low and require coarser resolution observations to enhance sensor signal to noise. Alternatively, radars illuminate the land surface directly and spatial resolution is driven by a desire to resolve individual landscape elements while maintaining frequent sampling and global coverage. Synthetic aperture radars (SARs) provide relatively fine-scale mapping of vegetation properties, though few provide both global coverage and frequent (<weekly) temporal sampling. Global data archives from other radars including SeaWinds and ASCAT (Advanced SCATterometer) meet these criteria but have relatively coarse (~ 25 –50 km) footprints. Post processing of overlapping image swaths from multiple satellite observations has produced enhanced resolution products at pixel scales from 4.5 km (SeaWinds) to 12.5 km (AMSR-E, SSM/I) (Early and Long, 2001). While resolution enhancement sacrifices some temporal fidelity, effective global coverage and temporal repeat of observations are generally sufficient for LSP monitoring.

Methods for LSP assessment

Methods for LSP assessment generally involve analyzing the shape of the VI seasonal curve to classify the timing of key thresholds or inflection points associated with vegetation life cycle events. Similar shape analysis or change detection methods are applicable to land parameter retrievals from both optical (VI, LAI) and microwave (e.g., VOD) remote sensing, while the majority of previous LSP studies and the following discussion focus on VI time series analysis. Analytical change detection approaches are susceptible to noise and associated high-frequency variability present in nearly all VI series. Methods for noise reduction include VI spatial compositing and temporal smoothing to reduce atmosphere cloud/aerosol contamination or other effects. Relatively simple classification approaches apply smoothed VI series with prescribed thresholds and backward-looking, moving window averages to distinguish LSP

events. An advantage of these methods is their relative simplicity. These methods work well for local scales and specific vegetation types but are difficult to implement globally because they do not account for heterogeneous surface or vegetation conditions, including multiple growth cycles. Efforts to circumvent these limitations include the use of globally variable, phenoregion-specific VI thresholds for assessing major LSP events and classifying phenotypic populations to reduce sensitivity to noise in individual pixels (White and Nemani, 2006).

Other classification techniques analyze VI series using mathematical models to identify life cycle thresholds. The MOD12Q2 algorithm analyzes smoothed EVI series using piecewise logistic functions to identify major LSP events coinciding with local extremes in temporal change rates of the VI seasonal cycle (Zhang et al., 2003). This methodology treats each pixel individually without the use of thresholds or empirical constants and is globally applicable. It is also capable of identifying multiple vegetation growth and senescence periods within a given year. A major assumption of these methods is that phenology can be characterized by a set of mathematical functions, which may not hold for areas that do not exhibit large seasonal changes in canopy biomass (e.g., evergreen forests) or where drought or defoliation events result in multiple VI curves with sharp peaks or broad plateaus. Also, the resulting LSP metrics can vary strongly depending on the particular data processing and classification methods used, even when based on the same satellite inputs (White et al., 2009).

Validation

LSP validation has involved comparisons against finer scale field measurements from native and cultivated indicator plants. These data predominantly represent species-specific events including timing of bud burst, flowering, and leaf drop from sparse in situ observation networks. Global phenology observations are expanding to include coordinated regional programs such as the USA National Phenology network (<http://www.usanpn.org/>). Direct comparisons between these data and LSP metrics are challenging because most field-based phenology observations involve individual species or local observations that may not be representative of landscape conditions within the surrounding satellite footprint.

Finer resolution satellite measurements from Landsat have been fit to sigmoid logistic growth models for spatial extrapolation of phenology observations to regional scales commensurate with global satellite remote sensing (Fisher et al., 2006). Tower eddy covariance measurements of land-atmosphere CO₂ exchange are more consistent with the scale of satellite remote sensing, facilitating more direct comparisons of field and satellite-based phenology metrics (Jones et al., 2012). Other approaches include the development of climate data-driven phenology models that incorporate information on individual plant species but provide a means for spatial extrapolation of sparse

field observations to the scale of satellite observations. An extension of this approach involves the integrated use of satellite observations, field measurements, and indicator clone models (Kathuroju et al., 2007). These methods provide the means for spatial scaling, prognostic modeling of future phenology events, and attribution of underlying processes influencing LSP patterns (White and Nemani, 2006).

Conclusion

Global operational LSP records have been generated from MODIS since 2000, with planned data continuity under the Visible Infrared Imaging Radiometer Suite (VIIRS) program. The AVHRR record continues to grow, enabling multi-decadal LSP studies that can distinguish impacts from weather variability, periodic climate oscillations (e.g., ENSO), and longer-term climate trends. Alternative global satellite data from ASCAT, AMSR-E, and other microwave sensors offer synergistic information on vegetation freeze/thaw status and canopy biomass structure that are insensitive to many of the environmental constraints on optical remote sensing. A wide array of algorithms exists for processing these data into LSP information (e.g., Zhang et al., 2003; Jönsson and Eklundh, 2004; White and Nemani, 2006). These various sensors and algorithms provide different perspectives, resolutions, and accuracies across the globe. Optical remote sensing delivers greater LSP precision where microwave data are too coarse to resolve subgrid-scale land cover dynamics. Alternatively, microwave remote sensing may deliver greater LSP accuracy where persistent clouds or reduced solar illumination limit optical retrievals. Together these data provide complimentary LSP information on vegetation greenness and biomass structure and insight regarding underlying environmental processes (Jones et al., 2012).

Future advancement of remote sensing science for LSP will likely incorporate data reduction and optimization techniques including empirical clustering and data assimilation for merging complimentary information from multiple sensors, land models, and observation networks to maximize LSP accuracy and understanding (e.g., Stöckli et al., 2011). These methods may vary for different applications, as agriculture may have different resolution and accuracy requirements than biosphere-climate monitoring. The development and extension of long-term global LSP datasets from reprocessing and reanalysis of long-lived sensor records will also allow more precise detection, analysis, and prediction of climate and land cover change effects. Improved capabilities for near-term phenology forecasting will involve the integration of satellite data with surface observations, weather prediction, and ecological models (White and Nemani, 2006) and benefit diverse applications including fire risk (Roads et al., 2005), human health (Lash et al., 2008), and crop yield predictions (Mkhabela et al., 2005).

Bibliography

- Early, D. S., and Long, D. G., 2001. Image reconstruction and enhanced resolution imaging from irregular samples. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 291–302.
- Fisher, J. I., Mustard, J. F., and Vadeboncoeur, M. A., 2006. Green leaf phenology at Landsat resolution: scaling from the field to the satellite. *Remote Sensing of Environment*, **100**, 265–279.
- Frolking, S., Milliman, T., McDonald, K., et al., 2006. Evaluation of the SeaWinds Scatterometer for regional monitoring of vegetation phenology. *Journal of Geophysical Research*, **111**, D17302.
- Jones, M. O., Kimball, J. S., Jones, L. A., and McDonald, K. C., 2012. Satellite passive microwave detection of North America start of season phenology. *Remote Sensing of Environment*, **123**, 324–333.
- Jönsson, P., and Eklundh, L., 2004. TIMESAT – a program for analyzing time-series of satellite sensor data. *Computers and Geosciences*, **30**, 833–845.
- Kathuroju, N., White, M. A., Symanzik, J., et al., 2007. On the use of the advanced very high resolution radiometer for development of prognostic land surface phenology models. *Ecological Modelling*, **201**, 144–156.
- Kimball, J. S., McDonald, K. C., Running, S. W., and Frolking, S. E., 2004. Satellite radar remote sensing of seasonal growing seasons for boreal and subalpine evergreen forests. *Remote Sensing of Environment*, **90**, 243–258.
- Lash, R. R., Brunsell, N. A., and Peterson, A. T., 2008. Spatiotemporal environmental triggers of Ebola and Marburg virus transmission. *Geocarto International*, **23**, 451–466.
- Min, Q., and Lin, B., 2006. Determination of spring onset and growing season leaf development using satellite measurements. *Remote Sensing of Environment*, **104**, 96–102.
- Mkhabela, M. S., Mkhabela, M. S., and Mashinini, N. N., 2005. Early maize yield forecasting in the four agro-ecological regions of Swaziland using NDVI data. *Agricultural and Forest Meteorology*, **129**, 1–9.
- Reed, B. C., White, M., and Brown, J. F., 2003. Remote sensing phenology. In Schwartz, M. D. (ed.), *Phenology: An Integrative Environmental Science*. Boston: Kluwer, pp. 365–382.
- Roads, J., Fujioka, F., Chen, S., and Burgan, R., 2005. Seasonal fire danger forecasts for the USA. *International Journal of Wildland Fire*, **14**, 1–18.
- Stöckli, R., Rutishauser, T., Baker, I., et al., 2011. A global reanalysis of vegetation phenology. *Journal of Geophysical Research*, **116**, G03020.
- Ulaby, F. T., Moore, R. K., and Fung, A. K., 1986. *Microwave Remote Sensing: Active and Passive*. Dedham: Artec House, Vol. 1–3.
- White, M. A., and Nemani, R. R., 2006. Real-time monitoring and short-term forecasting of land surface phenology. *Remote Sensing of Environment*, **104**, 43–49.
- White, M. A., de Beurs, K. M., Didan, K., et al., 2009. Intercomparison, interpretation and assessment of spring phenology in North America estimated from remote sensing for 1982 to 2006. *Global Change Biology*, **15**, 2335–2359.
- Zhang, X., Friedl, M. A., Schaaf, C. B., et al., 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, **84**, 471–475.

Cross-references

[Agriculture and Remote Sensing Forestry](#)

VOLCANISM

Michael J. Abrams

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

Volcanic eruptions

Definition

Volcanism: the set of geologic processes that result in the expulsion of lava, pyroclastics, and gases at the Earth's surface. The process by which magma and gases are transferred from Earth's interior to the surface.

Introduction

It is estimated that more than 500 million people live on or near active volcanoes around the world. The very eruptive products that present clear and frequent danger to the inhabitants are the primary reason for their presence: fertile volcanic soils. Currently, there are of the order of 1,000 volcanoes worldwide that are considered active, having erupted in the last 10,000 years (Simkin and Siebert, 1994). That is more than the number of active volcanologists, suggesting that remote sensing technology can substitute for human observers or ground-based instruments to keep an eye on active volcanoes over time. There are several volcanic phenomena that lend themselves to remote detection and monitoring: eruption plumes, laden with ash, SO₂, and water; low-temperature anomalies such as crater lakes and fumaroles; high-temperature anomalies, such as lava flows, pyroclastic flows, and lava domes; and deformation of the volcanic edifice.

Eruption cloud detection

The earliest reports of successful observations of volcanic eruptions showed eruption plumes captured by weather satellites, such as GOES (Geostationary Operational Environmental Satellite), GMS (Geostationary Meteorological Satellite), or AVHRR (Advanced Very High Resolution Radiometer). These instruments typically have spatial resolution of 1–15 km and repeat observation periods as short as 15 min. These satellites are currently used routinely by the Alaska Volcano Observatory (AVO) and the National Weather Service to detect and monitor elevated ground temperatures and eruption clouds at volcanoes in the North Pacific Region (Dean et al., 2002). Volcanoes in this region include those in Alaska, Kamchatka Peninsula, and the northern Kuril Islands. These data provide observations multiple times per day of over 100 active volcanoes in this region. Visible and infrared band data are automatically processed to generate images that optimize the detection of eruptions, including geometric and radiometric correction, and subsectioning to nine preassigned sectors. Each sector includes multiple



Volcanism, Figure 1 MODIS visible image of Chaiten volcano, Chile, captured May 31, 2008. The small red square is a thermal anomaly at the volcano's summit. The image covers about 500 km east–west.

volcanoes. Data are then processed to optimize detection of eruptions including single-band contrast, multiple-band combinations, and coastline overlays. JPEG and GIF images are generated for access from remote computers via the World Wide Web. The data are also automatically scanned to generate time-temperature graphs at specific volcanoes to detect thermal changes. Analysts routinely review the data and submit observation reports twice daily. During heightened alerts, the data are analyzed and observation reports are submitted at least four times per day. AVO reports eruptions through its communication network to government officials and the general public via voice, fax, and e-mail. All volcano-related observations are entered into a Web-accessible database.

Another satellite instrument that is used operationally to detect and monitor eruptions is MODIS (Moderate Resolution Imaging Spectroradiometer) on NASA's Terra and Aqua platforms. The MODIS Thermal Alert Web site was created as a real-time tool to monitor volcanic eruptions (Flynn et al., 2002). A review of some of the thermal anomalies associated with volcanic activity that have been discovered includes large-scale effusive eruptions like those at Nyamuragira, Africa and Kilauea, Hawaii, providing examples of the utility of monitoring hazards on a daily basis. Additional activity detected at Howard Island and Erta 'Ale may have otherwise gone unreported because of their inaccessibility. During a recent series of eruptions of Chaiten in Chile, MODIS captured many images, including one on May 31, 2008, shown in Figure 1. Monthly MODIS alert composites are used to summarize global volcanic activity. Currently, the MODIS Thermal Alert covers the globe four times per day from two satellites, each with a MODIS instrument.

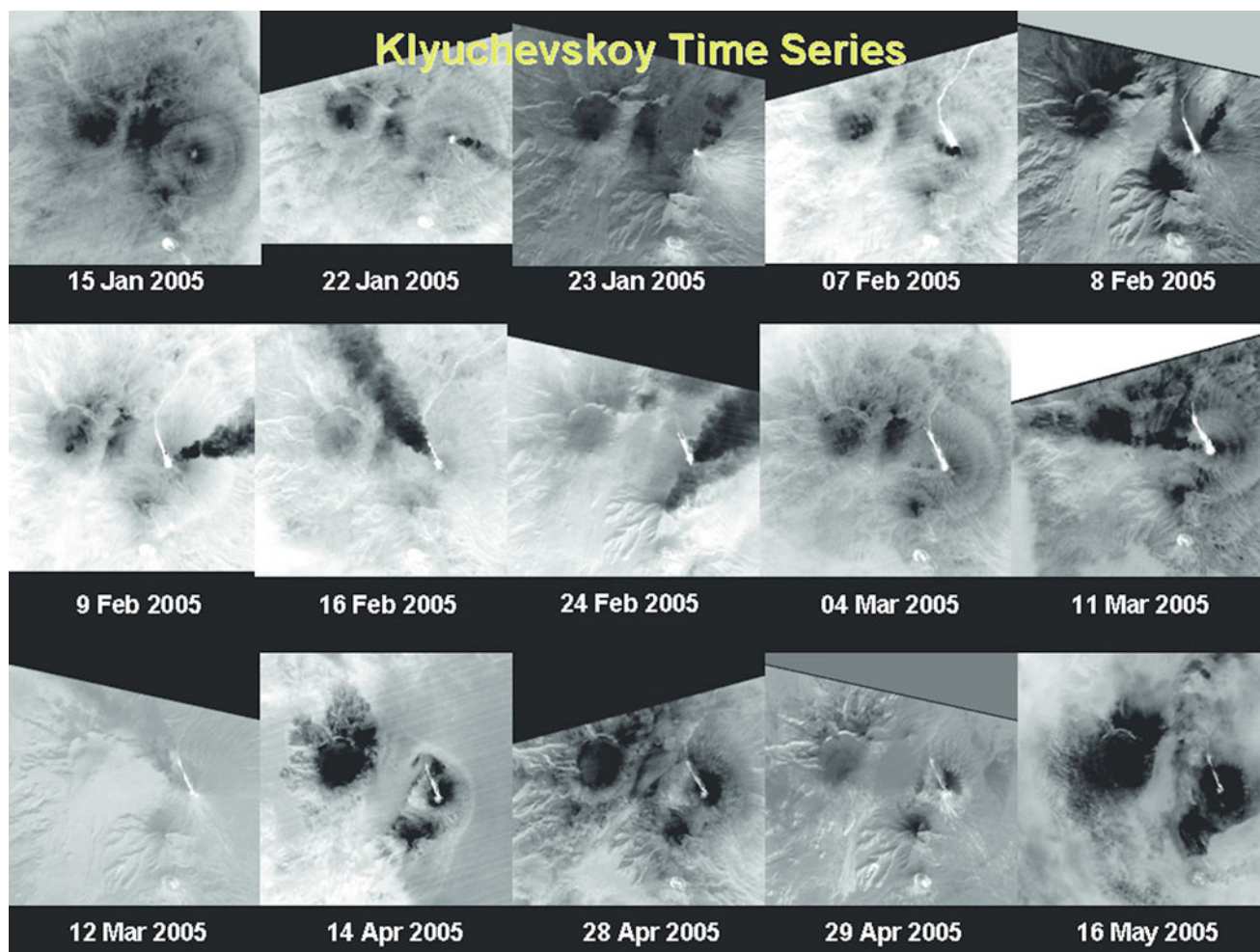
Thermal phenomena

Detecting and monitoring thermal phenomena associated with volcanoes, from space, requires sufficient spatial resolution to resolve the features, adequate wavelength regions to detect the thermal energy, and temporal coverage to be at "the right place at the right time." Volcanoes display precursory phenomena, as well as thermal features associated with eruptions. Included in the former are changes in temperature of lava lakes, growth of lava domes, and development or changes of temperature of fumaroles. Included in the latter are high-temperature or cooling lava flows, pyroclastic flows, debris flows, and dome growth.

Early work by Francis and Rothery (1987), Rothery et al. (1988), and Glaze et al. (1989) showed the potential use of Landsat Thematic Mapper data to estimate the temperature of volcanic phenomena, based on the presence of spectral bands in the visible, shortwave infrared, and thermal infrared wavelength regions, and spatial resolution of 30–120 m. A nice example of assembling a time history of thermal behavior as detected and measured by remote sensing, compared to eruptive history, was described by Oppenheimer et al. (1993). They assembled 15 Landsat images for 9 years from Lascar volcano, Chile, and documented the evolution of a lava dome within the summit crater. The thermal data indicated that Lascar experienced two cycles of lava dome activity. The first one culminated in an explosive eruption in 1986 that completely destroyed the dome.

Moderate resolution (<100 m) thermal infrared data are continuing to be used for monitoring volcanoes. A large experiment is ongoing over the North Pacific using low-resolution satellite data (MODIS, AVHRR) as triggers to schedule acquisitions of ASTER thermal data (90 m resolution) over erupting volcanoes (Ramsey and Dehn, 2005). A long time sequence has been acquired for Kliuchevskoi volcano in Kamchatka. Because of its location at high latitudes, polar-orbiting satellites have almost daily opportunities to obtain data there. From January 2005 to May 2005, 15 cloud-free images were acquired and are shown in Figure 2. The bright areas are hottest. Clearly evident is the early development of a summit crater thermal anomaly, either a lava dome or lava lake. In early February, the first of several pyroclastic flows were emplaced. Also quite evident are "cold" plumes coming out of the summit. They are colder than the ground because they have risen from the vent and cooled as they rose.

Wilson et al. (2007) reported on using satellite data to measure lava effusion rates. Kliuchevskoi volcano is the largest and one of the most active volcanoes on the Kamchatka Peninsula. The most recent eruption from January to August 2007 produced several strombolian eruptions, lava flows, mudflows, and ash plumes. To monitor the progress of the eruption, effusion rates were calculated from thermal infrared imagery. A volume of lava at a given temperature emits a fixed amount of thermal



Volcanism, Figure 2 ASTER thermal infrared images over Klyuchevskoi volcano, Kamchatka, Russia. Hot areas (lave dome, lava lake, and recently emplaced flows) are bright (hot). The eruption plumes are dark and cold. The time series was acquired from January to May 2005.

radiation measured at the satellite. By constraining temperature, volume with time (effusion rate) can be calculated. Effusion rates can give an indication of changes during an eruption, rather than after the activity has decreased or ceased. Effusion rates were consistently greater than $3.0 \text{ m}^3/\text{s}$ and reaching to $15.1 \text{ m}^3/\text{s}$. Calculations of effusion rates of Klyuchevskoi were conducted using a two-component analysis from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and compared to those made with coarser spatial resolution sensors: Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS). By constraining effusion rates in near real time, the potential hazards from the volcano can be effectively mitigated.

Deformation detection

Interferometric synthetic aperture radar, also abbreviated InSAR, is a radar technique that uses two or more SAR

images to generate maps of surface deformation, using differences in the phase of waves returning to the satellite. The technique can potentially measure centimeter-scale changes in deformation over time spans of days to years. InSAR can be used in a variety of volcanic settings, including deformation associated with eruptions, inter-eruption strain caused by changes in magma distribution at depth, gravitational spreading of volcanic edifices, and volcano-tectonic deformation signals. Early work on volcanic InSAR included studies on Mount Etna (Lundgren and Rosen, 2003) and Kilauea, with many more volcanoes being studied as the field developed (Amelung and Day, 2002; Pritchard and Simons, 2004). The technique is now widely used for academic research into volcanic deformation (Yun et al., 2007), although its use as an operational monitoring technique for volcano observatories has been limited by issues such as orbital repeat times, lack of archived data, coherence, and atmospheric errors.

Conclusion

The goal of remote sensing science is to contribute to predicting volcanic eruptions. As such, remote sensing provides several unique tools to the geophysicist's arsenal of observational capabilities. Presently, only very short-term predictions can be made, mainly based on in situ seismic data. But by increasing the number of volcanoes observed, building up time histories of thermal changes, and relating those to eruption behavior, it is hoped that one day potential eruptions can be successfully predicted, and damage to people and property mitigated.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- Amelung, F., and Day, S., 2002. InSAR observations of the 1995 Fogo, Cape Verde, eruption: implications for the effects of collapse events upon island volcanoes. *Geophysical Research Letters*, **29**, 1606–1611.
- Dean, K. G., Dehn, J., Engle, K., Izbekov, P., and Papp, K., 2002. Operational satellite monitoring of volcanoes at the Alaska volcano observatory. In Harris, A. J. H., Wooster, M., and Rothery, D. A., (eds.), *Monitoring Volcanic Hotspots Using Thermal Remote Sensing*, Advances in Environmental Monitoring and Modeling, Vol. 1, No. 3, pp. 70–97.
- Flynn, L., Wright, R., Garbeil, H., Harris, A., and Pilger, E., 2002. A global thermal alert system using MODIS. In Harris, A. J. H., Wooster, M., and Rothery, D. A., (eds.), *Monitoring Volcanic Hotspots Using Thermal Remote Sensing*, Advances in Environmental Monitoring and Modeling, Vol. 1, No. 3, pp. 37–69.
- Francis, P., and Rothery, D., 1987. Using the Landsat Thematic Mapper to detect and monitor active volcanoes: an example from Lascar Volcano, northern Chile. *Geology*, **15**, 614–617.
- Glaze, L., Francis, P., and Rothery, D., 1989. Measuring thermal budgets of active volcanoes by satellite remote sensing. *Nature*, **338**, 144–146.
- Lundgren, P., and Rosen, P., 2003. Source model for the 2001 flank eruption of Mt. Etna volcano. *Geophysical Research Letters*, **30**, 1388–1399.
- Oppenheimer, C., Francis, P., Rothery, D., Carlton, R., and Glaze, L., 1993. Infrared image-analysis of volcanic thermal features—Lascar Volcano, Chile, 1984–1992. *Journal of Volcanology and Geothermal Research*, **98**, 4269–4286.
- Pieri, D., and Abrams, M., 2004. ASTER watches the world's volcanoes: a new paradigm for volcanological observations from orbit. *Journal of Volcanology and Geothermal Research*, **135**, 13–28.
- Pritchard, M., and Simons, M., 2004. An InSAR-based survey of volcanic deformation in the southern Andes. *Geophysical Research Letters*, **15**, L15610–L15634.
- Ramsey, M., and Dehn, J., 2005. The changing behavior of silicic lava domes in Kamchatka: monitoring flow dynamics over four years with high resolution thermal infrared data. Abstracts of the General Assembly IAVCEI.
- Rothery, D., Francis, P., and Wood, C., 1988. Volcano monitoring using short wavelength infrared data from satellites. *Journal of Geophysical Research*, **93**, 7993–8008.
- Simkin, T., and Siebert, L., 1994. *Volcanoes of the World*, 2nd edn. Tucson/Washington, DC: Geoscience Press/The Smithsonian Institution.
- Wilson, R., Dehn, J., Prakash, A., and Wessels, R., 2007. Lava effusion rates from satellite remote sensing for Kliuchevskoi volcano, Kamchatka-Russia in 2007. American Geophysical Union, Abstract #V21A-0385.
- Yun, S., Zebker, H., and Segall, P., 2007. Interferogram formation in the presence of complex and large deformation. *Geophysical Research Letters*, **34**, L12305–L12308.

W

WATER AND ENERGY CYCLES

Taikan Oki¹ and Pat J.-F. Yeh²

¹Institute of Industrial Science, University of Tokyo,
Tokyo, Japan

²Department of Civil and Environmental Engineering,
National University of Singapore, Singapore, Singapore

Synonyms

Global water balance; Global water budget; Global water cycle

Definition

The global hydrological cycle can be described by the following physical processes which form a continuum of water movement. Complex pathways include the passage of water from the gaseous envelope around the planet called the atmosphere through the bodies of water on the surface of Earth such as the oceans, glaciers, and lakes and at the same time (or more slowly) passing through the soil and rock layers underground. Later, the water is returned to the atmosphere. A fundamental characteristic of the hydrological cycle is that it has no beginning and it has no end.

Introduction

The role of hydrological cycles in the Earth system, the amount of water on the Earth's surface, and its distribution in various reserves are first introduced together with the water cycles on the Earth. The concept of mean residence time, water stored in various parts over the Earth's surface as various phases, such as glacier, soil moisture, water vapor, and water flux among these reserves, such as precipitation, evaporation, transpiration, and runoff, is briefly explained with their roles in global climate system, and their quantitative estimates are presented. The zonally

averaged net transport of freshwater and the role of rivers in the global hydrological cycle are quantitatively shown. Finally, the measurements of certain fluxes and storages in the global hydrological cycle by using satellite remote sensing are reviewed.

Earth system and water

The Earth system is unique in that water exists in all three phases, that is, water vapor, liquid water, and solid ice, compared to the situations in other planets. The transport of water vapor is regarded as energy transport because of its large amount of latent heat exchange during phase change to liquid water (approximately 2.5×10^6 J kg⁻¹); therefore, water cycle is closely linked to energy cycle. Even though the energy cycle on the Earth is an "open system" driven by solar radiation, the amount of water on the Earth does not change on shorter than geological timescales (Oki, 1999; Oki et al., 2004), and the water cycle itself is a "closed system."

On the global scale, the hydrological cycles are associated with atmospheric circulation, which is driven by the unequal heating of the Earth's surface and atmosphere in latitude (Peixoto and Oort, 1992).

Annual mean absorbed solar energy at the top of the atmosphere is maximum near equator with approximately 300 W m⁻², decreases suddenly at higher latitudes, and is approximately 60 W m⁻² at Arctic and Antarctic regions. Emitted terrestrial radiative energy from the Earth at the top of the atmosphere is approximately 250 W m⁻² for $\pm 20^\circ$ north and south, gradually decreases at higher latitudes, and is approximately 175 W m⁻² at Arctic and 150 W m⁻² at Antarctic region. As a consequence, the net annual energy balance is positive (absorbing) for tropical and subtropical regions in $\pm 30^\circ$ north and south and negative in higher latitudes (Dingman, 2002).

If there are no atmospheric and oceanic circulation on the Earth, temperature difference on the Earth should have

Water and Energy Cycles, Table 1 World water reserves (simplified from Table 9 of “World water balance and water resources of the Earth” by UNESCO Korzun (1978). The last column, mean residence time, is from Table 34 of the report)

Form of water	Covering area (km ²)	Total volume (km ³)	Mean depth (m)	Share (%)	Mean residence time
World ocean	361,300,000	1,338,000,000	3,700	96.539	2,500 years
Glaciers and permanent snow cover	16,227,500	24,064,100	1,463	1.736	1,600 years
Groundwater	134,800,000	23,400,000	174	1.688	1,400 years
Ground ice in zones of permafrost strata	21,000,000	300,000	14	0.0216	10,000 years
Water in lakes	2,058,700	176,400	85.7	0.0127	17 years
Soil moisture	82,000,000	16,500	0.2	0.0012	1 years
Atmospheric water	510,000,000	12,900	0.025	0.0009	8 days
Marsh water	2,682,600	11,470	4.28	0.0008	5 years
Water in rivers	148,800,000	2,120	0.014	0.0002	16 days
Biological water	510,000,000	1,120	0.002	0.0001	A few hours
Artificial reservoirs		8,000			72 days
Total water reserves	510,000,000	1,385,984,610	2,718	100.00	

been more drastic; temperature in the equatorial zone should have been high enough that the outgoing terrestrial radiation balances the absorbed solar energy, and temperature in the polar regions, which should have been low enough, as well. In reality, there are atmospheric and oceanic circulations that lessen this expected temperature gradation in the absence of circulations.

Both atmosphere and ocean carry energy from the equatorial region toward both polar regions. In the case of atmosphere, the energy transport consists of sensible heat and latent heat fluxes (Masuda, 1988). The global water circulation is this latent heat transport itself, and water plays an active role in the atmospheric circulation; it is not a passive compound of the atmosphere, but it affects atmospheric circulation by both radiative transfer and latent heat release of phase change.

Water reserves, fluxes, and residence time

The total volume of water on the Earth is estimated as approximately $1.4 \times 10^{18} \text{ m}^3$, and it corresponds to a mass of $1.4 \times 10^{21} \text{ kg}$. Compared with the total mass of the Earth ($5.974 \times 10^{24} \text{ kg}$), the mass of water constitutes only 0.02 % of the planet, but it is critical for the survival of life on the Earth, and the Earth is called *Blue Planet* and *Living Planet*.

There are various forms of water on the Earth's surface. Approximately 70 % of its surface is covered with salt water, the oceans. Some of the remaining areas (continents) are covered by freshwater (lakes and rivers), solid water (ice and snow), and vegetation (which implies the existence of water). Even though the water content of the atmosphere is comparatively small (approximately 0.3 % by mass and 0.5 % by volume), approximately 60 % of the Earth is always covered by clouds (Rossow et al., 1993). The Earth is the planet whose surface is dominated by various phases of water.

Water on the Earth is stored in various reserves, and various water flows transport water from one to another. Water flow (mass or volume) per unit time is also called *water flux*.

The mean residence time in each reserve can be simply estimated from the total storage volume in the reserve and the mean flux rate to and from the reserve; there is even a distribution of flux rate coming in and going out from the storage (Chapman, 1972). The last column of Table 1 presents the values of the global mean residence time of water. Evidently, the water cycle on the Earth is a “stiff” differential system with the variability on many timescales, from a few weeks to thousands of years.

$$T_{\text{mean}} = \frac{\text{Total storage volume}}{\text{Mean flux rate}} \quad (1)$$

The mean residence time is also important to consider when water quality deterioration and restoration are discussed, since the mean residence time can be an index of how much water is turned over. Apparently, river water or surface water is more vulnerable than groundwater to be polluted; however, any measure to recover better water quality works faster for river water than groundwater. Since the major interests of hydrologists have been the assessment of volume, inflow, outflow, and chemical and isotopic composition of water, the estimation of the mean residence time of certain domains has been one of the major targets of hydrology.

Existence of water on the earth

Table 1 (simplified from a table in Korzun, 1978) introduces how much water is stored in which reserves on the Earth:

- The proportion in the ocean is large (96.5 %). Even though the classical hydrology has traditionally excluded ocean processes, the global hydrological cycle is never closed without including them. The ocean circulation carries huge amounts of energy and water. The surface ocean currents are driven by surface wind stress, and the atmosphere itself is sensitive to the [sea surface temperature](#). Temperature and salinity

determine the density of ocean water, and both factors contribute to the overturning and deep ocean general circulation.

- Other major reserves are solid water on the continent (glaciers and permanent snow cover) and groundwater. *Glacier* is the accumulation of ice of atmospheric origin generally moving slowly on land over a long period. Glacier forms discriminative U-shaped valley over land and remains moraine when it retreats. If a glacier “flows” into an ocean, the terminated end of the glacier often forms an iceberg. Glaciers react in comparatively longer timescale against climatic change, and they also induce isostatic responses of continental scale upheavals or subsidence in even longer timescale. Even though it is predicted that the thermal expansion of oceanic water dominates the anticipated *sea level rise* due to global warming, glaciers over land are also a major concern as the cause of sea level rise associated with global warming.
- *Groundwater* is the subsurface water occupying the saturated zone. It contributes to runoff in its low-flow regime, between floods. Deep groundwater may also reflect the long-term climatological situation. Groundwater in [Table 1](#) includes both gravitational and capillary water. Gravitational water is the water in the unsaturated zone (vadose zone), which moves under the influence of gravity. Capillary water is the water found in the soil above the water table by capillary action, a phenomenon associated with the surface tension of water in soils acting as porous media. Groundwater in Antarctica (roughly estimated as $2 \times 10^6 \text{ km}^3$) is excluded from [Table 1](#).
- *Soil moisture* is the water being held above the water table. It influences the energy balance at the land surface as a lack of available water suppresses evapotranspiration and as it changes surface albedo. Soil moisture also alters the fraction of precipitation partitioned into direct runoff and percolation. The water accounted for in the runoff cannot be evaporated from the same place, but the water infiltrated into soil may be uptaken by the capillary suction and evaporated again.
- The atmosphere carries *water vapor*, which influences the heat budget as latent heat. Condensation of water releases latent heat, heats up the atmosphere, and affects the atmospheric general circulation. *Liquid water in the atmosphere* is another result of condensation. *Clouds* significantly change the radiation in the atmosphere and at the Earth’s surface. However, as a volume, liquid (and solid) water contained in the atmosphere is quite little, and most of the water in the atmosphere exists as water vapor. *Precipitable water* is the total water vapor in the atmospheric column from land surface to the top of the atmosphere. Water vapor is also the major absorber in the atmosphere of both short-wave and longwave radiation.
- *Water in rivers* is very tiny as stored water all the time however, the recycling speed, which can be estimated as the inverse of the mean residence time, of river water

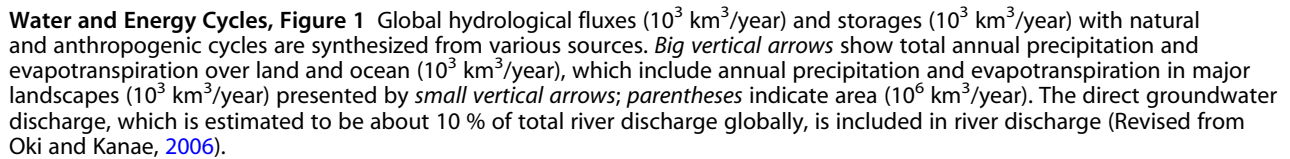
(river discharge) is relatively high, and it is important because most social applications ultimately depend on water as a renewable and sustainable resource.

Overall, the amount of water stored transiently in a soil layer, in the atmosphere, and in river channels is relatively minute, and the time spent through these subsystems is short, but, of course, they play dominant roles in the global hydrological cycle.

Water cycle on the earth

The water cycle plays many important roles in the climate system, and [Figure 1](#), revised from Oki and Kanae (2006), schematically illustrates various flow paths of water (Oki, 1999). Values are taken from [Table 1](#) and also calculated from the precipitation estimates by Xie and Arkin (1996). Precipitable water, water vapor transport, and its convergence are estimated using ECMWF objective analyses, obtained as 4 year mean from 1989 to 1992. The roles of these water fluxes in the global hydrological system are now briefly introduced:

- *Precipitation* is the water flux from atmosphere to land or *ocean* surface. It drives the hydrological cycle over land surface and changes surface salinity (and temperature) over the ocean and affects its thermohaline circulation. *Rainfall* refers to the liquid phase of precipitation. Part of it is intercepted by canopy over vegetated areas, and the remaining part reaches the Earth’s surface as throughfall. Highly variable, intermittent, and concentrated behavior of precipitation in time and space domain compared to other major hydrological fluxes mentioned below makes the observation of this quantity and the aggregation of the process complex and difficult.
- *Snow* has special characteristics compared with rainfall. Snow may be accumulated, the albedo of snow is quite high (as high as clouds), and the surface temperature will not rise above 0°C until the completion of snowmelt. Consequently, the existence of snow changes the surface energy and water budget enormously. A snow surface typically reduces the aerodynamic roughness, so that it may also have a dynamical effect on the atmospheric circulation and hydrological cycle.
- *Evaporation* is the return flow of water from the surface to the atmosphere and takes the latent heat flux from the surface. The amount of evaporation is determined by both atmospheric and hydrological conditions. From the atmospheric point of view, the fraction of incoming solar energy to the surface leading to latent and sensible heat flux is important. Wetness at the surface influences this fraction because the ratio of actual evapotranspiration to the potential evaporation is reduced due to drying stress. The stress is sometimes formulated as a resistance, and such a condition of evaporation is classified as hydrology driven. If the land surface is wet enough compared to the available energy for evaporation, the condition is classified as atmosphere driven.



		N.P.	S.P.	N. At.	S. At.	Indian	Arctic	Inner	Total
From rivers	Asia	4.7	0.4	0.2		3.3	2.7	0.1	11.4
	Europe			1.7		0.0	0.7		2.4
	Africa			-0.2	0.9	-0.2		-0.4	0.1
	N. America	2.9		4.8			1.1		8.8
	S. America	0.5	0.4	5.7	8.3				14.9
	Australia		0.1			0.1			0.2
From atmosphere	Antarctica		1.0		0.1	0.8			1.9
	Total	8.1	1.9	12.2	9.3	4.0	4.5	-0.3	39.7
	$-\nabla_H \cdot \overline{Q}$	9.9	-11.1	-12.7	-14.0	-14.0	2.2		-39.7
	Grand total	18.0	-9.2	-0.5	-4.7	-10.0	6.7	-0.3	0.0

- Transpiration is the evaporation of water through *stomata* of leaves. It has two special characteristics different from evaporation from soil surfaces. One is that the resistance of stomata is related not only to the dryness of soil moisture but also to the physiological conditions of the vegetation through the opening and closing of stomata. Another is that roots can transfer water from deeper soil than in the case of evaporation from bare soil.

Vegetation also modifies surface energy and water balance by altering surface albedo and by intercepting precipitation and evaporating this rainwater.

- Runoff at the hillslope scale is nonlinear and a complex process. Surface runoff could be generated when rainfall or snowmelt intensity exceeds the infiltration rate of the soil or precipitation falls over saturated land surface. Saturation at land surface can be formed mostly by topographic concentration mechanism along hillslopes. Infiltrated water in the upper part of the hillslope flows down the slope and discharges at the bottom of the hillslope. Because of the highly variable heterogeneity of topography, soil properties such as conductivity and porosity, and precipitation, basic equations such as Richards' equation, which can express the runoff process fairly well at a point scale or hillslope scale, cannot be directly applied to the macroscale because of its nonlinearity.

Runoff returns water to the ocean which may have been transported inland in vapor phase by atmospheric advection. The runoff into oceans is also important for the freshwater balance and the salinity of the oceans. Rivers carry not only water mass but also sediments, chemicals, and various nutritional matters from continents to seas. Without rivers, the global hydrological cycles on the Earth will never close.

The global water cycle unifies these components consisting of the state variables (precipitable water, soil moisture, etc.) and the fluxes (precipitation, evaporation, etc.).

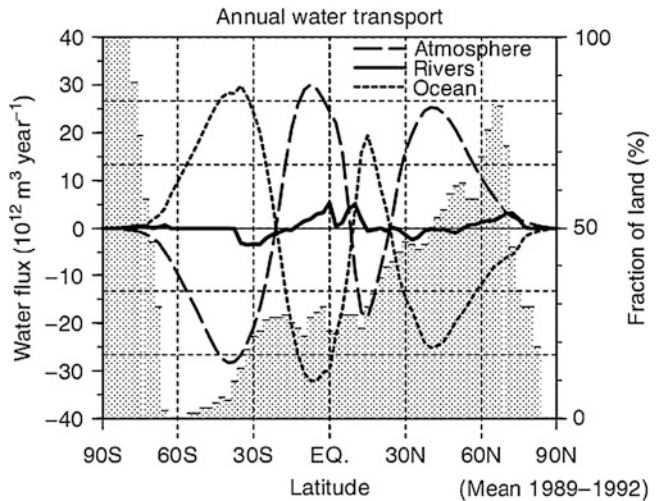
Zonally averaged net transport of freshwater

The meridional (north–south direction) distribution of the zonally averaged annual *energy* transports by the atmosphere and the oceans has been evaluated, even though there are quantitative problems in estimating such values (Trenberth and Solomon, 1994). However, the corresponding distribution of water transport has not often been studied although the cycles of energy and water are closely related. Wijnffels et al. (1992) used values of the convergence of water vapor flux in the atmosphere (Q) from Bryan and Oort (1984) and discharge data from Baumgartner and Reichel (1975) to estimate the freshwater transport by oceans and atmosphere, but their results seem to have large uncertainties, and they did not present the freshwater transport by rivers.

The annual freshwater transport in the meridional (north–south) direction can be estimated from Q and river discharge with the geographical information such as the location of river mouths and basin boundaries (Oki et al., 1995). Results are introduced in the next section.

Rivers in global hydrological cycle

The freshwater supply to the ocean has an important effect on the thermohaline circulation because it changes the salinity and thus the density. The impacts of freshwater supply to ocean are enhanced in the case of large river basins because they concentrate freshwater from large



Water and Energy Cycles, Figure 2 The annual freshwater transport in the meridional (north–south) direction by atmosphere, ocean, and rivers (land) (Oki et al., 1995). Water vapor flux transport of $20 \times 10^{12} \text{ m}^3/\text{year}$ corresponds to approximately $1.6 \times 10^{15} \text{ W}$ of latent heat transport. Shaded bars behind the lines indicate the fraction of land at each latitudinal belt.

areas to their river mouths. It also controls the formation of sea ice and its temporal and spatial variations. Annual freshwater transport by rivers and the atmosphere to each ocean is summarized in Table 2 based on the atmospheric water balance (Oki, 1999). Some part of the water vapor flux convergence remains in the inland basins. There are a few negative values in Table 2, suggesting that net freshwater transport occurs from the ocean to the continents. This is physically impossible and is caused by errors in the source data. Although detailed discussion of the values in Table 2 may not be meaningful, it is nevertheless interesting that such an analysis does make at least qualitative sense using the atmospheric water balance method with the geographical information on basin boundaries and the location of river mouths. In this analysis, it should be noted that the total amount of freshwater transport into the oceans from the surrounding continents has the same order of magnitude as the freshwater supply that comes directly from the atmosphere, expressed by Q .

The annual freshwater transport in the meridional direction has also been estimated based on the atmospheric water balance with the results shown in Figure 2. The estimates in Figure 2 are the net transport, that is, in the case of oceans, it is the residual of northward and southward freshwater fluxes by all ocean currents globally, and it cannot be compared directly with individual ocean currents such as the Kuroshio and the Gulf Stream. It should be noted that the directions of river flows are mostly steady unlike the ocean or atmospheric circulations and concentrate the freshwater in one direction throughout the year.

Transports by the atmosphere and by the ocean have almost the same absolute values at each latitude but with different signs. The transport by rivers is about 10 % of these other fluxes globally (this may be an underestimation because Q tends to be smaller than the river discharge observed at a land surface). The negative (southward) peak by rivers at 30°S is mainly due to the Paraná River in South America, and the peaks at the equator and 10 °N are due to rivers in South America, such as the Magdalena and Orinoco. Large Russian rivers, such as the Ob, Yenisey, and Lena, carry freshwater toward the north between 50 °N and 70 °N.

These results suggest that the hydrological processes over land play nonnegligible roles in the climate system, not only by the exchange of energy and water at the land surface but also through the transport of freshwater by rivers, which affects the water balance of the oceans and forms a part of the hydrological circulation on the Earth among the atmosphere, continents, and oceans.

Remote sensing in global hydrological cycle

Over the past 20 years, significant progress has been made toward routine monitoring of certain fluxes and storages in the global hydrological cycle by satellite remote sensing, while continued progress is anticipated from upcoming missions. Table 3 gives an overview on the past, current, and planned future hydrological remote sensing

capabilities. Some of these important achievements in measuring global water cycle components are discussed below:

- *Precipitation* – The remote sensing precipitation observation systems were originated about three decades ago (Griffith et al., 1978). Most rainfall products measured from spaceborne platforms, for example, NEXRAD (Next-Generation Radar) and TRMM (Tropical Rainfall Measuring Mission) Precipitation Radar rainfall measurements, are typically provided at large space–time scales suitable for coarse-scale meteorological applications, such as climatologic analysis and water balance studies. Satellite rainfall retrieval is subject to errors caused by various factors ranging from infrequent sampling to the high complexity and variability in the relationship of the measurement to precipitation parameters. Therefore, advanced retrieval algorithm and spatial downscaling techniques (e.g., PERSSIAN, Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks; Sorooshian et al., 2000) are necessary to be applied to the satellite rainfall data for the purpose of the study of global water cycle. The planned Global Precipitation Measurement (GPM; <http://gpm.gsfc.nasa.gov/>), which is an international satellite mission scheduled to launch in 2014, envisions a large constellation of passive microwave sensors to provide global rainfall products at the temporal resolution of ~3 h and spatial resolution

Water and Energy Cycles, Table 3 Summary of past, current, and planned future hydrological satellite remote sensing capabilities

Satellite sensors/missions	Hydrological variables measured	Time period of observation
Geostationary Operational Environmental Satellite (GOES)	Precipitation	1978–present
Special Sensor Microwave/Imager (SSM/I)	Precipitation, snow water equivalent, snow extent	1987–present
Scanning Multichannel Microwave Radiometer (SMMR)	Snow water equivalent, snow extent	1978–1987
European Remote Sensing-1 (ERS-1) Radar Altimeter and SAR	Surface water height, soil moisture	1991–2000
TOPEX/Poseidon	Surface water height	1993–2005
European Remote Sensing-2 (ERS-2) Radar Altimeter and SAR	Surface water height, soil moisture	1996–present
Tropical Rainfall Measuring Mission (TRMM)	Precipitation	1998–present
Advanced Microwave Sounding Unit (AMSU)	Precipitation	1998–present
Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra	Snow extent, evapotranspiration ^a	2000–present
Geosat Follow-On Mission	Surface water height	2000–present
Atmospheric Infrared Sounder (AIRS)/Aqua	Water vapor	2002–present
Advanced Microwave Scanning Radiometer – EOS (AMSR-E)/Aqua	Soil moisture, snow water equivalent	2002–present
Moderate Resolution Imaging Spectroradiometer (MODIS)/Aqua	Snow extent, evapotranspiration ^a	2002–present
ENVISAT Radar Altimeter-2 and ASAR	Surface water height, soil moisture	2002–present
Jason-1	Surface water height	2002–present
Gravity Recovery and Climate Experiment (GRACE)	Total water storage, soil moisture, ^a groundwater ^a	2002–present
Soil Moisture and Ocean Salinity (SMOS)	Soil moisture	2009 - present
Soil Moisture active Passive (SMAP)	Soil moisture	Scheduled in 2014
Global Precipitation Measurement (GPM)	Precipitation	Scheduled in 2014

^aNot directly measured; ancillary data required in the estimation

of $\sim 100 \text{ km}^2$ (Hossain and Lettenmaier, 2006; Smith et al., 2007). The GPM mission will provide almost real-time rainfall information at three hourly sampling interval on a global basis, thus allowing hydrologists an opportunity to improve flood prediction capability for medium to large river basins, especially in the underdeveloped world where in situ precipitation gauge networks are sparse. Another current community-wise agenda on the satellite precipitation missions is the “Program for Evaluation of High Resolution Precipitation Products (PEHRPP),” which is an effort led by the International Precipitation Working Group (IPWG) to evaluate the quality of currently available high-resolution satellite rainfall products (Ebert et al., 2007).

- *Terrestrial water storage* – Another contribution of remote sensing technologies to understand the global hydrological cycle has been the measurement of terrestrial water storage (TWS) and its components (Famiglietti, 2004). The major achievements in this aspect include (1) soil moisture retrieval (Jackson et al., 2002; Njoku et al., 2003) from the Soil Moisture and Ocean Salinity (SMOS; Pellarin et al., 2003) and Hydrosphere Satellite Mission (Hydros; Entekhabi et al., 2004), (2) surface water height measurement using the altimetry (Alsdorf and Lettenmaier, 2003; Alsdorf et al., 2007), and (3) integrated measurement of TWS from the Gravity Recovery and Climate Experiment (GRACE) mission (Tapley et al., 2004), among others. TWS is a fundamental component of the global water cycle and an integrated measure of surface and subsurface water availability (i.e., the sum of soil moisture, groundwater, snow and ice, waters in vegetation and biomass, and surface water in lakes, reservoirs, wetlands, and river channels snow). It has great importance for the management of water resources, agriculture, and ecosystem health. TWS controls the partitioning of precipitation into evaporation and runoff and the partitioning of net radiation into the sensible and latent heat fluxes, with significant implications for precipitation recycling, hydrological extremes (i.e., flood and drought), and land memory processes (Shukla and Mintz, 1982; Eltahir and Bras, 1996; Eltahir and Yeh, 1999; Koster et al., 2004), while TWS change is a basic quantity in closing the terrestrial water balance from local to regional and global scales (Ngo-Duc et al., 2005; Güntner et al., 2007; Yeh and Famiglietti, 2008). Despite its importance, its role in the global hydrological cycle has received little attention relative to other hydrological processes (Lettenmaier and Famiglietti, 2006), and there are no extensive networks currently in existence for monitoring TWS changes. Satellite observations of the Earth’s time-variable gravity field from the GRACE mission (Tapley et al., 2004) present a new opportunity to explore the feasibility of monitoring TWS variations from space. Short-term (monthly, seasonal, and interannual) temporal variations in gravity on land are largely due to corresponding changes in vertically

integrated terrestrial water storage (Wahr et al., 2004). This has allowed for the first time observations of variations in total TWS at large river basins (Swenson et al., 2003; Chen et al., 2005; Seo et al., 2006; Winsemius et al., 2006) to continental scales (Wahr et al., 2004; Ramillien et al., 2005; Schmidt et al., 2006; Klees et al., 2007; Syed et al., 2008); for new approaches to remote estimation of discharge (Syed et al., 2005) and evapotranspiration (Rodell et al., 2004); of groundwater variations (Rodell and Famiglietti, 2002; Yeh et al., 2006) and snow water storage (Frappart et al., 2005); and for validation and improvement of the terrestrial water balance in the global land surface hydrological models (Niu and Yang, 2006; Swenson and Milly, 2006).

- *Water vapor* – Another example of remote sensing measurements is the global water vapor distribution. Water vapor is one of the most important greenhouse gases; small amounts of water vapor in the form of clouds can strongly affect both shortwave and longwave radiations. Buoyancy created by changes in the phase of water largely drives the vertical motion of the atmosphere. The Atmospheric Infrared Sounder (AIRS) on board NASA’s Earth Observing System (EOS) *Aqua* spacecraft measures water vapor at $\sim 2 \text{ km}$ vertical resolution with 10–15 % accuracy in clear sky conditions (Moustafa et al., 2006). Such information is available only in a small percentage of the globe because most of the AIRS pixels are cloud contaminated. Success in applying AIRS data has been achieved via selectively choosing cloud-free pixels. Additionally, surface lidar can also provide useful information on water vapor, temperature, and winds in clear air below clouds although they are rather expensive and delicate instruments of limited deployment and coverage. Global Positioning System (GPS) also holds promise to be used as a water vapor sensor when combined with independent temperature analyses.

Summary

The global hydrological cycle consisting of oceans, water vapor in the atmosphere, and terrestrial water, is essential to the Earth system. The cycle is closed by the exchange of water and energy fluxes between these reservoirs. Although the amounts of water in the atmosphere and river channels are relatively small, their fluxes are large, and hence a critical role in society, which is dependent on water as a renewable resource. The ultimate goal of the hydrological science is to enhance the understanding of global water and energy cycles on various spatial and temporal scales through monitoring and modeling, and the outcomes should also be beneficial and accessible to other scientific disciplines, the general public, and the decision makers.

Bibliography

- Alsdorf, D. E., and Lettenmaier, D. P., 2003. Tracking fresh water from space. *Science*, **301**, 1491–1494.

- Alsdorf, D., Bates, P., Melack, J., Wilson, M., and Dunne, T., 2007. Spatial and temporal complexity of the Amazon flood measured from space. *Geophysical Research Letters*, **34**, L08402, doi:10.1029/2007GL029447.
- Baumgartner, F., and Reichel, E., 1975. *The World Water Balance: Mean Annual Global, Continental and Maritime Precipitation, Evaporation and Runoff*. Munchen: Ordenbourg, p. 179.
- Bryan, F., and Oort, A., 1984. Seasonal variation of the global water balance based on aerological data. *Journal of Geophysical Research*, **89**(11), 717–730.
- Chapman, T. G., 1972. *Estimating the Frequency Distribution of Hydrologic Residence Time*. IAHS-UNESCO-WMO, Vol. 1, pp. 136–152.
- Chen, J., Rodell, M., Wilson, C. R., and Famiglietti, J. S., 2005. Low degree spherical harmonic influences on Gravity Recovery and Climate Experiment (GRACE) water storage estimates. *Geophysical Research Letters*, **32**, L14405, doi:10.1029/2005GL022964.
- Dingman, S. L., 2002. *Physical Hydrology*, 2nd edn. Upper Saddle River: Prentice-Hall, p. 646.
- Ebert, E., Janowiak, J. E., and Kidd, C., 2007. Comparison of near real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society*, **88**, 47–64.
- Eltahir, E. A. B., and Bras, R. L., 1996. Precipitation recycling. *Reviews of Geophysics*, **34**(3), 367–378.
- Eltahir, E. A. B., and Yeh, P. J.-F., 1999. On the asymmetric response of aquifer water level to droughts and floods in Illinois. *Water Resources Research*, **35**(4), 1199–1217.
- Entekhabi, D. D., et al., 2004. The hydrosphere state (hydros) satellite mission: an Earth system pathfinder for global mapping of soil moisture and land freeze. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, 2184–2195.
- Famiglietti, J. S., 2004. Remote sensing of terrestrial water storage, soil moisture and surface waters. In Sparks, R. S. J. and Hawkesworth, C. J. (eds.), *The State of the Planet: Frontiers and Challenges in Geophysics, Geophysical Monograph Series*. Washington, DC: American Geophysical Union, Vol. 150, pp. 197–207.
- Frappart, F., Ramillien, G., Biancamaria, S., Mognard, N., and Cazenave, A., 2005. Evolution of high-latitude snow mass derived from the GRACE gravimetry mission (2002–2004). *Geophysical Research Letters*, **32**, doi:10.1029/2005GL024778.
- Griffith, G. C., Woodley, W. L., and Grube, P. G., 1978. Rain estimation from geosynchronous satellite imagery-visible and infrared studies. *Monthly Weather Review*, **106**, 1153–1171.
- Güntner, A., Stuck, J., Werth, S., Döll, P., Verzano, K., and Merz, B., 2007. A global analysis of temporal and spatial variations in continental water storage. *Water Resources Research*, **43**, W05416, doi:10.1029/2006WR005247.
- Hossain, F., and Lettenmaier, D. P., 2006. Flood prediction in the future: recognizing hydrologic issues in anticipation of the Global Precipitation Measurement mission. *Water Resources Research*, **42**, W11301, doi:10.1029/2006WR005202.
- Jackson, T. J., Gasiewski, A. J., Oldak, A., Klein, M., Njoku, E. G., Yevgrafov, A., Christiani, S., and Bindlish, R., 2002. Soil moisture retrieval using the C-band polarimetric scanning radiometer during the Southern Great Plains 1999 Experiment. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, 2151–2161.
- Klees, R., Zapreeva, E. A., Winsemius, H. C., and Savenije, H. H. G., 2007. The bias in GRACE estimates of continental water storage variations. *Hydrology and Earth System Sciences*, **11**, 1227–1241.
- Korzun, V. I., 1978. World water balance and water resources of the earth. In *Studies and Reports in Hydrology*. Paris: UNESCO, Vol. 25.
- Koster, R. D., et al., 2004. Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138–1140, doi:10.1126/science.1100217.
- Lettenmaier, D. P., and Famiglietti, J. S., 2006. Water from on high. *Nature*, **444**, 562–563.
- Masuda, K., 1988. Meridional heat transport by the atmosphere and the ocean; analysis of FGGE data. *Tellus*, **40A**, 285–302.
- Moustafa, T., et al., 2006. AIRS: improving weather forecasting and providing new data on greenhouse gases. *Bulletin of the American Meteorological Society*, **87**, 911–926.
- Ngo-Duc, T., Laval, K., Polcher, J., and Cazenave, A., 2005. Contribution of continental water to sea level variations during the 1997–1998 El Niño Southern Oscillation event: comparison between Atmospheric Model Intercomparison Project simulations and TOPEX/Poseidon satellite data. *Journal of Geophysical Research*, **110**, D09103, doi:10.1029/2004JD004940.
- Niu, G.-Y., and Yang, Z. L., 2006. Assessing a land surface model's improvements with GRACE estimates. *Geophysical Research Letters*, **33**, L07401, doi:10.1029/2005GL025555.
- Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K., and Nghiem, S. V., 2003. Soil moisture retrieval from AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing*, **41**, 215–229.
- Oki, T., 1999. The global water cycle. In Browning, K., and Gurney, R. (eds.), *Global Energy and Water Cycles*. Cambridge: Cambridge University Press, pp. 10–27.
- Oki, T., and Kanae, S., 2006. Global hydrological cycles and world water resources. *Science*, **313**(5790), 1068–1072.
- Oki, T., Musiak, K., Matsuyama, H., and Masuda, K., 1995. Global atmospheric water balance and runoff from large river basins. *Hydrological Processes*, **9**, 655–678.
- Oki, T., Entekhabi, D., and Harrold, T., 2004. The global water cycle. In Sparks, R., and Hawkesworth, C. (eds.), *State of the Planet: Frontiers and Challenges in Geophysics*. Washington, DC: AGU Publication. Geophysical Monograph Series, Vol. 150, p. 414.
- Peixoto, J. P., and Oort, A. H., 1992. *Physics of Climate*. New York: American Institute of Physics, p. 520.
- Pellarin, T., Wigneron, J.-P., Calvet, J.-C., and Waldteufel, P., 2003. Global soil moisture retrieval from a synthetic L-band brightness-temperature data set. *Journal of Geophysical Research*, **108**(D12), 4364, doi:10.1029/2002JD003086.
- Ramillien, G., Frappart, F., Cazenave, A., and Güntner, A., 2005. Time variations of land water storage from an inversion of 2 years of GRACE geoids. *Earth and Planetary Science Letters*, **235**, 283–301.
- Rodell, M., and Famiglietti, J. S. 2002. The potential for satellite-based monitoring of groundwater storage changes using GRACE: the high plains aquifer, Central U. S. *Journal of Hydrology*, **263**, 245–256.
- Rodell, M., et al., 2004. The global land data assimilation system. *Bulletin of the American Meteorological Society*, **85**, 381–394.
- Rossow, W. B., Walker, A. W., and Garder, L. C., 1993. Comparison of ISCCP and other cloud amounts. *Journal of Climate*, **6**, 2394–2418.
- Schmidt, R., et al., 2006. GRACE observations of changes in continental water storage. *Global and Planetary Change*, **50**, 112–126, doi:10.1016/j.gloplacha.2004.11.018.
- Seo, K.-W., Wilson, C. R., Famiglietti, J. S., Chen, J. L., and Rodell, M., 2006. Terrestrial water mass load changes from Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, **42**, W05417, doi:10.1029/2005WR004255.
- Shukla, J., and Mintz, Y., 1982. Influence of land-surface evapotranspiration on the Earth's climate. *Science*, **215**, 1498–1501.

- Smith, E., et al., 2007. The international global precipitation measurement (GPM) program and mission: an overview. In Levizzani, V., and Turk, F. J. (eds.), *Measuring Precipitation from Space: EURAINSAT and the Future*. Dordrecht: Kluwer.
- Sorooshian, S., Hsu, K. L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, B., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of American Meteorology Society*, **81**, 2035–2046.
- Swenson, S., and Milly, P. C. D. 2006. Climate model biases in seasonality of continental water storage revealed by satellite gravimetry. *Water Resources Research*, **42**, W03201, doi:10.1029/2005WR004628.
- Swenson, S., Wahr, J., and Milly, P. C. D., 2003. Estimated accuracies of regional water storage variations inferred from the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, **39**(8), 1223, doi:10.2002WR001808.
- Swenson, S. C., Yeh, P. J.-F., Wahr, J., and Famiglietti, J. S., 2006. A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. *Geophysical Research Letters*, **33**, L16401, doi:10.1029/2006GL026962.
- Syed, T. H., Famiglietti, J. S., Chen, J., Rodell, M., Seneviratne, S. I., Viterbo, P., and Wilson, C. R., 2005. Total basin discharge for the Amazon and Mississippi river basins from GRACE and a land-atmosphere water balance. *Geophysical Research Letters*, **32**, L24404, doi:10.1029/2005GL024851.
- Syed, T. H., Famiglietti J. S., Rodell, M., Chen, J., and Wilson C. R. 2008. Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research*, **44**, W02433, doi:10.1029/2006WR005779.
- Tapley, B., Bettapur, S., Watkins, M., and Reigber, C., 2004. The gravity recovery and climate experiment: mission overview and first results. *Geophysical Research Letters*, **31**, L09607, doi:10.1029/2004GL019920.
- Trenberth, K. E., and Solomon, A., 1994. The global heat balance: heat transports in the atmosphere and ocean. *Climate Dynamics*, **10**, 107–134.
- Wahr, J., Swenson, S., Zlotnicki, V., and Velicogna, I., 2004. Time variable gravity from GRACE: first results. *Geophysical Research Letters*, **31**, L11501, doi:10.1029/2004GL019779.
- Wijffels, S. E., Schmitt, R. W., Bryden, H. L., and Stigebrandt, A., 1992. Transport of freshwater by the oceans. *Journal of Physical Oceanography*, **22**, 155–162.
- Winsemius, H. C., Savenije, H. H. G., van de Giesen, N. C., van den Hurk, B. J. J. M., Zapreeva, E. A., and Klees, R., 2006. Assessment of Gravity Recovery and Climate Experiment (GRACE) temporal signature over upper Zambezi. *Water Resources Research*, **42**, W12201, doi:10.1029/2006WR005192.
- Xie, P., and Arkin, P. A., 1996. Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *Journal of Climate*, **9**, 840–858.
- Yeh, P. J.-F., and Famiglietti, J., 2008. Regional terrestrial water storage change and evapotranspiration from terrestrial and atmospheric water balance computations. *Journal of Geophysical Research-Atmospheres*, **113**, doi:10.1029/2007JD009045.
- Yeh, P. J.-F., Famiglietti, J., Swenson, S. C., and Rodell, M., 2006. Remote sensing of groundwater storage changes using gravity recovery and climate experiment (GRACE). *Water Resources Research*, **42**, W12203, doi:10.1029/2006WR005374.

Cross-references

[Cloud Properties](#)
[Ice Sheets and Ice Volume](#)
[Rainfall](#)
[Terrestrial Snow](#)
[Water Vapor](#)

WATER RESOURCES

Taikan Oki¹ and Pat J.-F. Yeh²

¹Institute of Industrial Science, University of Tokyo, Tokyo, Japan

²Department of Civil and Environmental Engineering, National University of Singapore, Singapore, Singapore

Synonyms

Freshwater resources; Runoff

Definition

Water resources mainly correspond to freshwater which can be utilized by human beings for irrigation, industrial purposes, and domestic uses. Even though the stocks of water in natural and artificial reservoirs are helpful to increase the available water resources for human society, the flow of water should be mainly regarded as water resources since water is a naturally circulating resource that is constantly recharged. Maximum (potential) availability of renewable freshwater resources under given climatic condition has been traditionally regarded as precipitation minus evapotranspiration, which corresponds to runoff. However, transpiration from soil moisture through crops is contributing to human society and the flux is regarded as water resources as well. Therefore, evapotranspiration from soil moisture in croplands is called as green water nowadays, and conventional water resources withdrawn from rivers, surface water, and groundwater are called as blue water.

Introduction

All organisms, including humans, require water for their survival. Therefore, ensuring adequate and sufficient water supplies is essential for human well-being. Stored waters, such as deep groundwater and water in reservoirs, are often considered as water resources. However, they are a part of hydrological cycles, and groundwater is also circulating even though its speed could be rather slow. Some groundwater is called “fossil water,” which implies that the aquifer was recharged long term ago and that it will be depleted if being overly exploited just as fossil fuels. From this point of view, flows of water should be considered as water resources in the assessments, designing, and planning of sustainable water usage.

Global water balance and water resources

Conventionally, available freshwater resources are defined as annual runoff estimated by compiling observed river discharge data or by using water balance equation (i.e., the residual of annual precipitation minus evapotranspiration; see Baumgartner and Reichel, 1975; Korzun, 1978). Such an approach has and is continuing to provide information on the annual freshwater resources for many countries. Theoretically, available annual freshwater resources can be derived using annual

precipitation (P) and evapotranspiration (E) measurements; however, the residual term annual runoff (R), namely, annual available freshwater resources, is relatively small compared to P and E . P and E are approximately 1,100 and 1,200 mm/year, respectively, over the ocean and 800 and 500 mm/year over the land (Oki, 1999). Therefore, the estimation of R could contain certain amount of uncertainties due to even comparatively small errors in the estimation of P and E .

Atmospheric water balance computation using the information of water vapor flux convergence could be alternatively used to estimate global runoff distribution owing to the advent of four-dimensional data assimilation (4DDA) technique in atmospheric science (Oki et al., 1995). Even though the microwave remote sensing of precipitable water (vertically integrated water vapor) and temperature profiles is contributing to improve the quality of 4DDA data, this approach is suitable for global overview of the distribution of available freshwater resources (Trenberth et al., 2007).

Model-based estimation of available freshwater resources

Relatively simple water balance models have been used to estimate grid-based available freshwater resources in the world (Alcamo et al., 2000; Vörösmarty et al., 2000). Later, land surface models (LSMs) were applied for the estimation of global water cycles (Oki et al., 2001; Dirmeyer et al., 2006) and for the global water resources assessment by comparing the demand side for both the twentieth century and the future (Shen et al., 2008). Some of these estimates on global water balance were calibrated by multiplying an empirical factor inferred from available observed river discharge data. However, recent model developments are capable to estimate runoff with adequate accuracy without the need of calibration (Hanasaki et al., 2008a). Such estimates by using numerical models require external atmospheric forcing data such as precipitation and temperature, and the accuracy of estimated freshwater resources is highly dependent on the quality of the forcing data (Oki et al., 1999). Satellite remote sensing can provide reasonable estimates of the forcing data for the regions with low density of in situ observations, particularly with regard to the quantities such as precipitation and downward shortwave and longwave radiation. Additionally, 4DDA reanalysis products have been commonly used as the forcing data for air temperature, humidity, and wind speed. In order to assure the quality of the estimations, the results of land surface simulations have been validated, typically, by comparing with in situ discharge observations. However, recently, various satellite remote sensing observations have been used to validate hydrological variables other than discharge such as inundated area (Prigent et al., 2007) and total terrestrial water storage (Kim et al., 2009). Also, some remote sensing variables, such as top-soil moisture (Reichle and Koster, 2005) and snow cover

(Zaitchik and Rodell, 2009), have been assimilated into modeling system to reduce the simulation uncertainty.

Global distribution of available freshwater resources

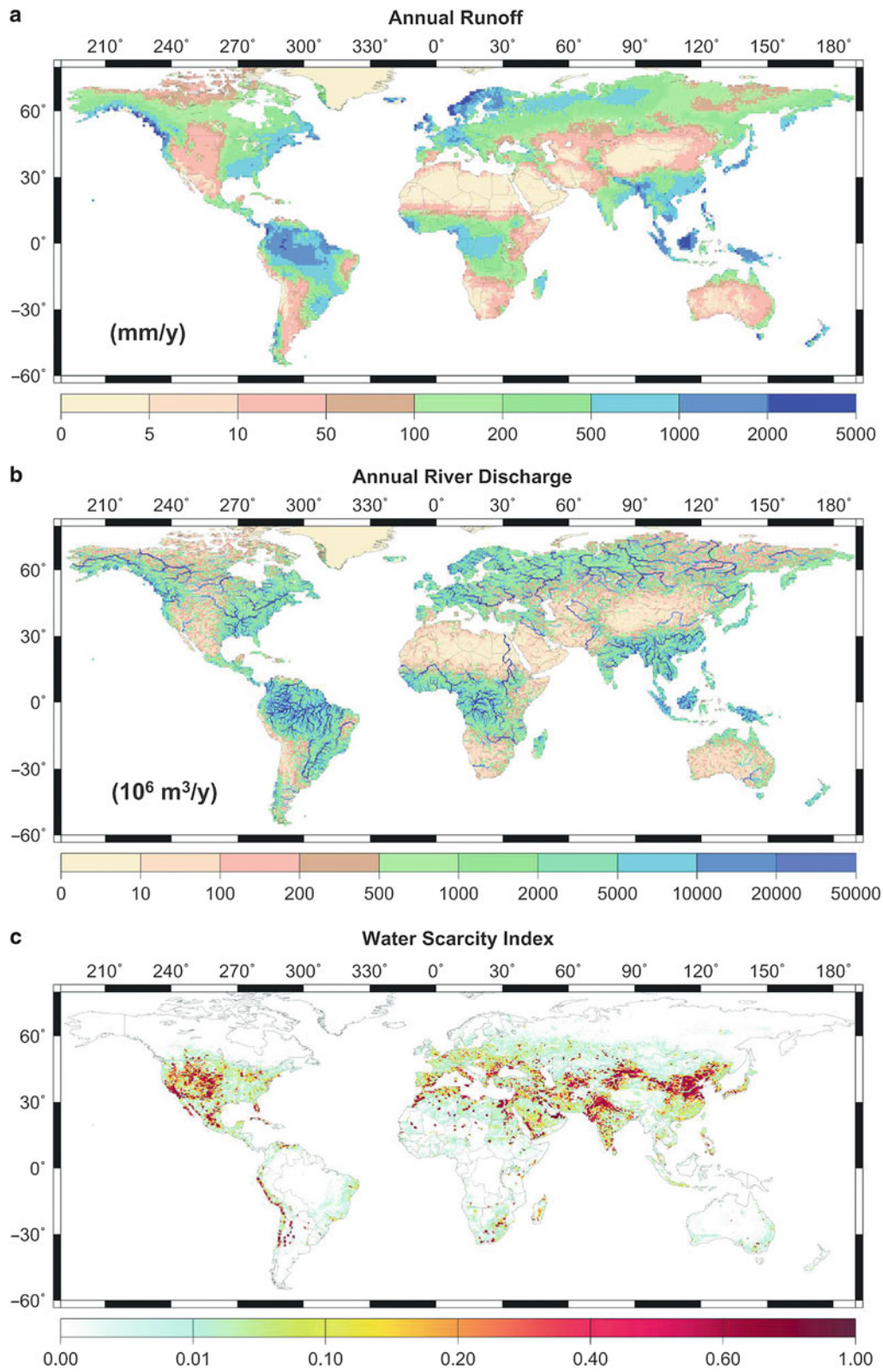
Figure 1 illustrates the global distribution of annual runoff estimated by land surface models from the multi-model ensemble simulations under the Global Soil Wetness Project (Dirmeyer et al., 1999). Annual runoff (Figure 1) can be considered as the maximum available renewable freshwater resources (RFWR) if waters from upstream cannot be reused at downstream due to consumptive use or water pollution (Oki et al., 2001). Runoff is accumulated through river channels and realized as river discharge (Figure 1). River discharge can be considered as the potentially maximum available RFWR if all the water from upstream can be used. Both runoff and river discharge are concentrated in limited areas, and their amounts range from nearly zero in desert areas to more than 2,000 mm/year in the tropics and greater than 200,000 m³/s of discharge on average near the river mouth of the Amazon (Oki and Kanae, 2006).

Some macroscale hydrological models recently developed for water resources assessments have been equipped with a reservoir operation scheme (e.g., Haddeland et al., 2006; Hanasaki et al., 2006) in order to simulate the “real” hydrological cycles and provide information on actually available freshwater resources. These water resources have been significantly influenced by anthropogenic activities and modified from the “natural” hydrological cycles even on the global scale in “Anthropocene” (Crutzen, 2002).

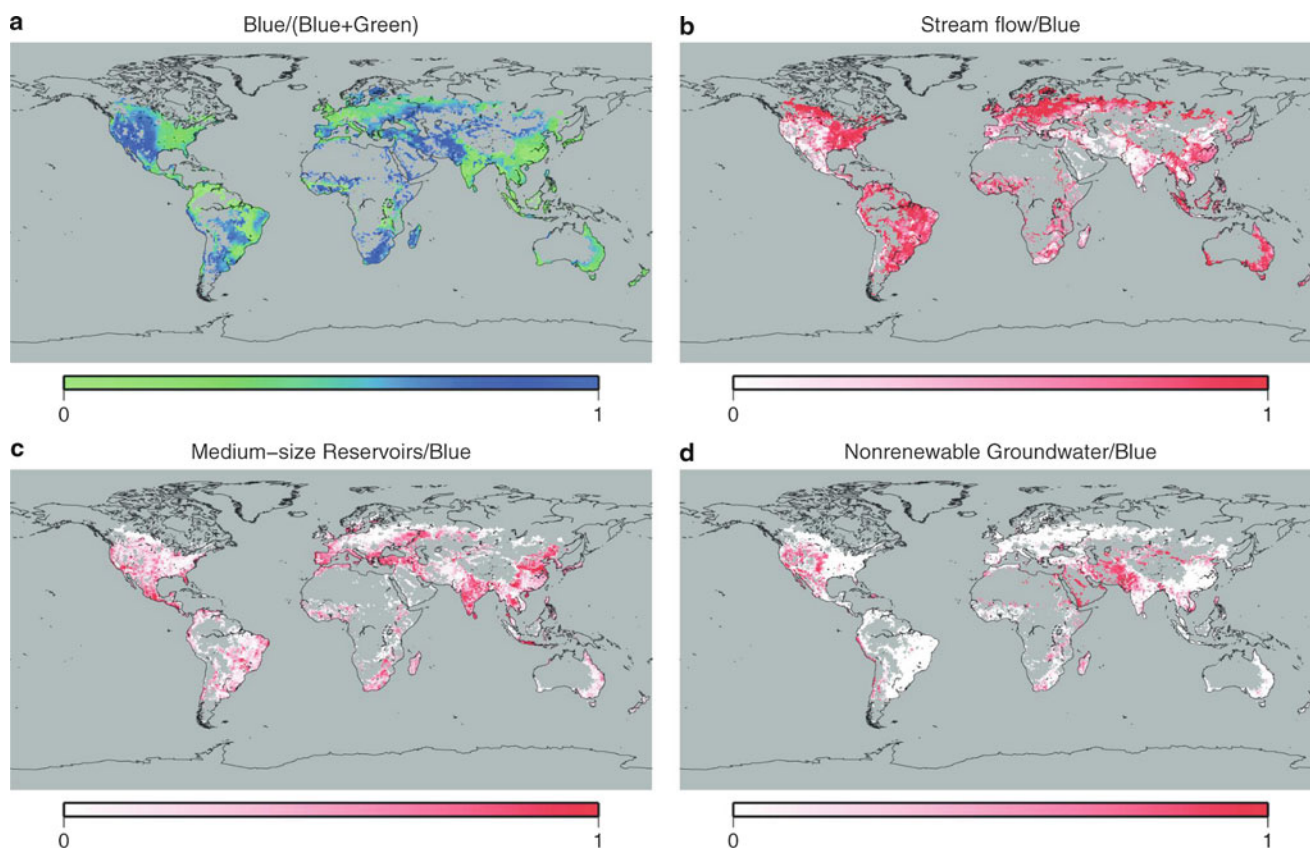
Water resources for crop growth

An integrated water resources model can further be linked to a crop growth sub-model designed for inferring the timing and quantity of irrigation requirement and estimating environmental flow (Hanasaki et al., 2008a). Such an approach enables the assessment of the balances between demand and supply of water resources on a daily time scale. Using this approach (Hanasaki et al., 2008a), a gap in the sub-annual distribution of water availability and water use can be detected in the Sahel, the Asian monsoon region, and southern Africa, where the conventional water scarcity indices such as the ratio of annual water withdrawal to water availability or the available annual water resources per capita (Falkenmark and Rockström, 2004) cannot properly detect the stringent balance between water demand and supply (Hanasaki et al., 2008b).

Moreover, macroscale numerical models can be associated with a scheme tracing the origin of flow path as if tracing the isotopic ratio of water (Yoshimura et al., 2004). Such a water flow-tracing function, if incorporated into an integrated water resources model (e.g., Hanasaki et al., 2008a) with the consideration of multiple sources of water withdrawals including streamflow, medium-size



Water Resources, Figure 1 Global distribution of (a) mean annual runoff (mm/year), (b) mean annual discharge (million m^3/year), and (c) water scarcity index R_{ws} . Water stress is higher for regions with larger R_{ws} (Oki and Kanae, 2006).



Water Resources, Figure 2 (a) The ratio of blue water to the total evapotranspiration during a cropping period from irrigated cropland (the total of green and blue water). The ratios of (b) streamflow, (c) medium-size reservoirs, and (d) nonrenewable groundwater withdrawals to blue water (Hanasaki et al., 2010).

reservoirs, and nonrenewable groundwater in addition to precipitation on the croplands, is able to trace the origin of water used to produce the major crops (Hanasaki et al., 2010).

Figure 2a illustrates the ratio of blue water to the total evapotranspiration during the cropping period in irrigated croplands. Here the blue (green) water is defined as the amount of water evapotranspiration originated from irrigation (precipitation) (see Falkenmark and Rockström, 2004). Figure 2a shows distinctive geographical distribution of the pattern of the dependence on blue water. Total annual blue water consumption is estimated approximately as 1,500 km³/year, which is about 20 % of the total consumptive use of approximately 7,000 km³/year of water resources in croplands during the cropping period (Hanasaki et al., 2009). Further, the ratios of the source of blue water are shown for streamflow including the influence of large reservoirs, medium-size reservoirs, and nonrenewable groundwater in Figure 2b–d, respectively. Areas highly dependent on nonrenewable groundwater are detected in the Pakistan, Bangladesh, and in western part of India, north and western parts of China, some regions in the Arabian Peninsula, and the western

part of the United States and Mexico. Cumulative nonrenewable groundwater withdrawals estimated by the model correspond fairly well with the country statistics of total groundwater withdrawals, and such an integrated model has the ability to quantify global virtual water flow (Allan, 1998; Oki and Kanae, 2004) and “water footprint” (Hoekstra and Chapagain, 2007) through the major crop water consumption (Hanasaki et al., 2009).

Remote sensing applications in water resources

The last two decades have witnessed significant achievements toward routine monitoring of global hydrologic cycle components by using remote sensing techniques, while continued progress is anticipated from upcoming missions. Among these space efforts, the following missions are particularly relevant to water resources applications:

1. Integrated measurement of terrestrial water storage (TWS) from the Gravity Recovery and Climate Experiment (GRACE) mission (Tapley et al., 2004), launched in 2002
2. Soil moisture retrieval from the Soil Moisture and Ocean Salinity (SMOS; Kerr et al., 2000), launched

by the European Space Agency (ESA) in 2009, and the Soil Moisture Active and Passive (SMAP) mission (Entekhabi et al., 2004), planned to launch by NASA in 2014

3. Surface water height measurement using the altimetry from the Surface Water Ocean Topography (SWOT) mission (Alsdorf and Lettenmaier, 2003), planned to launch by NASA in 2014

TWS is a fundamental component of in closing the terrestrial water balance from local to regional and global scales. As an integrated measure of surface and subsurface water availability, TWS bears significant implications for water resources planning and management. Despite its importance, there are no extensive networks currently in existence for monitoring TWS changes. Satellite observations of Earth's time-variable gravity field from the GRACE mission present a new opportunity to explore the feasibility of monitoring TWS variations from space. Short-term (monthly, seasonal, and interannual) temporal variations in gravity on land are largely due to corresponding changes in vertically integrated terrestrial water storage (Wahr et al., 2004). This has allowed for the first time observations of variations in total TWS at large river basins (Swenson et al., 2003) to continental scales (Wahr et al., 2004). Also, application using GRACE data has been made in the estimation of discharge (Syed et al., 2005), evapotranspiration (Rodell et al., 2004), groundwater variations (Yeh et al., 2006), snow water storage (Frappart et al., 2005), surface water dynamics (Han et al., 2009), lateral redistribution of water storage through river networks (Kim et al., 2009), and validation and improvement of global land surface hydrological models (Niu and Yang, 2006).

The SMOS and upcoming SMAP missions are the first dedicated satellite missions to measure surface soil moisture levels globally. Soil moisture is an important factor which interfaces water and energy exchanges between the land surface/atmosphere and is the most important hydrological quantity for agriculture. Water management for irrigation is a critical issue for global crop production and food safety. Root zone soil moisture, which is strongly related with transpiration (green water), will be more reliably estimated by merging SMOS and SMAP observations with a land surface model in a data assimilation system, even though the enhanced technologies in those satellites are still limited to directly observe only topsoil moisture. It will enable meteorology and food agencies to forecast crop yield and enhance the capabilities of crop water stress decision support systems, monitor global climate change, detect droughts, and conduct flood forecasting and weather prediction.

The SWOT satellite mission and its wide-swath (20–120 km) altimetry technology for repeated elevation measurements can measure the water height variations of the global oceans and terrestrial surface waters accurately. For water resources applications, hydrological observations of the temporal and spatial variations in water

volumes stored in all wetlands, lakes, and reservoirs are extremely important. However, because of coarse spatial (>100 km) and temporal (>1 month) resolutions, previous researches using altimetry (e.g., Birkett et al., 2002) have been used in only limited conditions and objectives. By measuring water height and area variations in higher spatial (2 m × 10 m to 2 m × 60 m) and temporal (2 weeks) resolutions which allow accurate estimation of river discharges and lake/wetland storages in remote regions, SWOT will contribute to a fundamental understanding of the terrestrial branch of the global water cycle and hence benefit global water resource planning and management.

Conclusions

Current advancement of remote sensing technology has not proved to be well capable of observing global water fluxes; thus, it is not an easy task to accurately estimate available freshwater resources based on remotely sensed data. However, remote sensing technique can help in providing necessary information such as meteorological forcing data, land use and land cover, extent of surface water bodies, and topography for the modeling estimation of water resources. Remote sensing of cropland coverage, planting date, and harvesting date, and detection of irrigated areas are necessary for assessing the balances between demand and supply of water resources in addition to the social information, such as population distribution, urban area, industrial water use, and domestic water use. Remote sensing of hydrological quantities, such as snow cover, soil moisture, and temporal change of total terrestrial water storage, can provide great values to constrain and validate hydrological model simulations and thereby assure accurate estimates of freshwater resources from model simulations of water storages and fluxes.

Bibliography

- Alcamo, J., Henrichs, T., and Rösch, T., 2000. *World Water in 2025 – Global Modeling and Scenario Analysis for the World Commission on Water for the 21st Century, Technical Report*. Kassel: Centre for Environmental Systems Research, University of Kassel.
- Allan, J. A., 1998. Virtual water: a strategic resource. *Global solution to regional deficits*. *Groundwater*, **36**(4), 545–546.
- Alsdorf, D. E., and Lettenmaier, D. P., 2003. Tracking fresh water from space. *Science*, **301**, 1491–1494.
- Baumgartner, F., and Reichel, E., 1975. *The World Water Balance: Mean Annual Global, Continental and Maritime Precipitation, Evaporation and Runoff*. Munchen: Ordenbourg, p. 179.
- Birkett, C. M., Mertes, L. A. K., Dunne, T., Costa, M. H., and Jasinski, M. J., 2002. Surface water dynamics in the Amazon basin: application of satellite radar altimetry. *Journal of Geophysical Research*, **107**(D20), 8059, doi:10.1029/2001JD000609.
- Crutzen, P. J., 2002. Geology of mankind-the Anthropocene. *Nature*, **415**, 23.
- Dirmeyer, P. A., Dolman, A. J., and Sato, N., 1999. The pilot phase of the global soil wetness project. *Bulletin of the American Meteorological Society*, **80**, 851–878.

- Dirmeyer, P. A., Gao, X. A., Zhao, M., Guo, Z. C., Oki, T., and Hanasaki, N., 2006. GSWP-2 multimodel analysis and implications for our perception of the land surface. *Bulletin of the American Meteorological Society*, **87**, 1381–1397.
- Entekhabi, D., Njoku, E., Houser, P., Spencer, M., Doiron, T., Smith, J., Girard, R., Belair, S., Crow, W., Jackson, T., Kerr, Y., Kimball, J., Koster, R., McDonald, K., O'Neill, P., Pultz, T., Running, S., Shi, J. C., Wood, E., and van Zyl, J., 2004. The Hydrosphere State (HYDROS) mission concept: an Earth system pathfinder for global mapping of soil moisture and land freeze/thaw. *IEEE Geoscience and Remote Sensing*, **42**(10), 2184–2195.
- Falkenmark, M., and Rockström, J., 2004. *Balancing Water for Humans and Nature*. London: Earthscan, p. 247.
- Falloon, P. D., and Betts, R. A., 2006. The impact of climate change on global river flow in HadGEM1 simulations. *Atmospheric Science Letters*, **7**, 62–68.
- Frappart, F., Ramillien, G., Biancamaria, S., Mognard, N., and Cazenave, A., 2005. Evolution of high-latitude snow mass derived from the GRACE gravimetry mission (2002–2004). *Geophysical Research Letters*, **32**, doi:10.1029/2005GL024778.
- Haddeland, I., Lettenmaier, D. P., and Skaugen, T., 2006. Effects of irrigation on the water and energy balances of the Colorado and Mekong river basins. *Journal of Hydrology*, **324**, 210–223.
- Han, S.-C., Kim, H., Yeo, I.-Y., Yeh, P. J.-F., Oki, T., Seo, K.-W., Alsdorf, D., and Luthcke, S. B., 2009. Dynamics of surface water storage in the Amazon inferred from measurements of inter-satellite distance change. *Geophysical Research Letters*, **36**, L09403, doi:10.1029/2009GL037910.
- Hanasaki, N., Kanae, S., and Oki, T., 2006. A reservoir operation scheme for global river routing models. *Journal of Hydrology*, **327**, 22–41.
- Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., and Tanaka, K., 2008a. An integrated model for the assessment of global water resources -Part 1: Model description and input meteorological forcing. *Hydrology and Earth System Sciences*, **12**, 1007–1025.
- Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., and Tanaka, K., 2008b. An integrated model for the assessment of global water resources -Part 2: Applications and assessments. *Hydrology and Earth System Sciences*, **12**, 1027–1037.
- Hanasaki, N., Inuzuka, T., Kanae, S., Oki, T., 2010. An estimation of global virtual water flow and sources of water withdrawal for major crops and livestock products using a global hydrological model. *Journal of Hydrology*, **384**, 232–244. doi:10.1016/j.jhydrol.2009.09.028
- Hirabayashi, Y., and Kanae, S., 2009. First estimate of the future global population at risk of flooding. *Hydrological Research Letters*, **3**, 6–9.
- Hoekstra, A. Y., and Chapagain, A. K., 2007. Water footprints of nations: water use by people as a function of their consumption pattern. *Water Resources Management*, **21**, 35–48.
- Kerr, Y., Font, J., Waldteufel, P., and Berger, M., 2000. The second of ESA's opportunity missions: the soil moisture and ocean salinity mission – SMOS. *ESA Earth Observation Quarterly*, **66**, 18f.
- Kim, H., Yeh, P., Oki, T., and Kanae, S., 2009. Role of rivers in the seasonal variations of terrestrial water storage over global basins. *Geophysical Research Letters*, **36**, L17402, doi:10.1029/2009GL039006.
- Korzun, V. I., 1978. *World Water Balance and Water Resources of the Earth*. Paris: UNESCO. Studies and Reports in Hydrology, Vol. 25.
- Miller, J. R., Russell, G. L., and Caliri, G., 1994. Continental-scale river flow in climate models. *Journal of Climate*, **7**, 914–928.
- Ngo-Duc, T., Oki, T., and Kanae, S., 2007. A variable streamflow velocity method for global river routing model: model description and preliminary results. *Hydrology and Earth System Sciences Discussions*, **4**, 4389–4414.
- Niu, G.-Y., and Yang, Z. L., 2006. Assessing a land surface model's improvements with GRACE estimates. *Geophysical Research Letters*, **33**, L07401, doi:10.1029/2005GL025555.
- Oki, T., 1999. The global water cycle. In Browning, K., and Gurney, R. (eds.), *Global Energy and Water Cycles*. Cambridge/New York: Cambridge University Press, pp. 10–27.
- Oki, T., and Kanae, S., 2004. Virtual water trade and world water resources. *Water Science and Technology*, **49**(7), 203–209.
- Oki, T., and Kanae, S., 2006. Global hydrological cycles and world water resources. *Science*, **313**(5790), 1068–1072.
- Oki, T., and Sud, Y. C., 1998. Design of total runoff integrating pathways (TRIP) a global river channel network. *Earth Interactions*, **2**, 1–37.
- Oki, T., Musiak, K., Matsuyama, H., and Masuda, K., 1995. Global atmospheric water balance and runoff from large river basins. *Hydrological Processes*, **9**, 655–678.
- Oki, T., Kanae, S., and Musiak, K., 1996. River routing in the global water cycle, *GEWEX News*, **6**, WCRP, International GEWEX Project Office, 4–5.
- Oki, T., Nishimura, T., and Dirmeyer, P., 1999. Assessment of annual runoff from land surface models using total runoff integrating pathways (TRIP). *Journal of the Meteorological Society of Japan*, **77**, 235–255.
- Oki, T., Agata, Y., Kanae, S., Saruhashi, T., Yang, D. W., and Musiak, K., 2001. Global assessment of current water resources using total runoff integrating pathways. *Hydrological Sciences Journal*, **46**, 983–995.
- Oki, T., Entekhabi, D., and Harrold, T., 2004. The global water cycle. In Sparks, R., and Hawkesworth, C. (eds.), *State of the Planet: Frontiers and Challenges in Geophysics*. Washington, DC: AGU. Geophysical Monograph Series, Vol. 150, p. 414.
- Prigent, C., Papa, F., Aires, F., Rossow, W. B., and Matthew, E., 2007. Global inundation dynamics inferred from multiple satellite observations, 1993–2000. *Journal of Geophysical Research*, **112**, D12107, doi:10.1029/2006JD007847.
- Reichle, R., and Koster, R., 2005. Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. *Geophysical Research Letters*, **32**, L02404, doi:10.1029/2004GL021700.
- Rodell, M., et al., 2004. The global land data assimilation system. *Bulletin of the American Meteorological Society*, **85**, 381–394.
- Shen, Y., Oki, T., Utsumi, N., Kanae, S., and Hanasaki, N., 2008. Projection of future world water resources under SRES scenarios: water withdrawal. *Hydrological Sciences Journal*, **53**(1), 11–33.
- Swenson, S., Wahr, J., and Milly, P. C. D., 2003. Estimated accuracies of regional water storage variations inferred from the gravity recovery and climate experiment (GRACE). *Water Resources Research*, **39**(8), 1223, doi:10.2002WR001808.
- Syed, T. H., Famiglietti, J. S., Chen, J., Rodell, M., Seneviratne, S. I., Viterbo, P., and Wilson, C. R., 2005. Total basin discharge for the Amazon and Mississippi river basins from GRACE and a land-atmosphere water balance. *Geophysical Research Letters*, **32**, L24404, doi:10.1029/2005GL024851.
- Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., and Watkins, M. M., 2004. GRACE measurements of mass variability in the Earth system. *Science*, **305**(5683), 503–505.
- Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J., 2007. Estimates of the global water budget and its annual cycle using observational and model data. *Journal of Hydrometeorology*, **8**, 758–769.

- Vörösmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B., 2000. Global water resources: vulnerability from climate change and population growth. *Science*, **289**, 284–288.
- Wahr, J., Swenson, S., Zlotnicki, V., and Velicogna, I., 2004. Time variable gravity from GRACE: first results. *Geophysical Research Letters*, **31**, L11501, doi:10.1029/2004GL019779.
- Yeh, P. J.-F., Famiglietti, J., Swenson, S. C., and Rodell, M., 2006. Remote sensing of groundwater storage changes using gravity recovery and climate experiment (GRACE). *Water Resources Research*, **42**, W12203, doi:10.1029/2006WR005374.
- Yoshimura, K., Oki, T., Ohte, N., and Kanae, S., 2004. Colored moisture analysis estimates of variations in 1998 Asian monsoon water sources. *Journal of the Meteorological Society of Japan*, **82**, 1315–1329.
- Zaitchik, B., and Rodell, M., 2009. Forward-looking assimilation of MODIS-derived snow-covered area into a land surface model. *Journal of Hydrometeorology*, **10**, 130–148.

Cross-references

[Agriculture and Remote Sensing](#)
[Crop Stress](#)
[Earth Radiation Budget, Top-of-Atmosphere Radiation](#)
[Irrigation Management](#)
[Rainfall](#)
[Surface Water](#)
[Snowfall](#)
[Soil Moisture](#)
[Water and Energy Cycles](#)

WATER VAPOR

Eric Fetzer
 Jet Propulsion Laboratory, California Institute of
 Technology, Pasadena, CA, USA

Synonyms

Atmospheric humidity; Atmospheric moisture

Definition

Water vapor. Water in gaseous form, usually mixed with dry air in the Earth's atmosphere.

Water vapor mixing ratio. The ratio of density of water vapor to the density of air. Typical values range from a few grams/kilogram in the tropical lower troposphere to a few micrograms/kilogram around the tropopause.

Relative humidity. The ratio of water vapor partial pressure to *saturation vapor pressure*. A common measure of water vapor amount, but not conserved as an air parcel changes temperature. *Mixing ratio* is conserved.

Saturation vapor pressure. Maximum partial pressure of water vapor adjacent to a plain surface of water or ice. Water vapor saturation vapor pressure follows the *Clausius-Clapeyron relation* for water and varies approximately exponentially with temperature.

Clausius-Clapeyron relation. Equation relating equilibrium pressure of a gaseous substance adjacent to a solid or liquid surface. The Clausius-Clapeyron relation varies exponentially with temperature at about 7 %/K near

freezing. In terrestrial atmospheric sciences, this term generally refers to water vapor with respect to liquid water or ice.

Latent heat of vaporization. Heat required to convert a unit mass of liquid water to water vapor. The same heat is released when water vapor condenses to liquid, as happens in clouds. Analogous latent heat of fusion refers to conversion between ice and liquid. Water vapor latent heats are very large: about 2.3×10^6 J/kg for conversion of liquid water to vapor compared to dry air heat capacity of 717 J/kg-K.

See American Meteorological Society (2010) and Wallace and Hobbs (2006) for additional terms, including dew point, *frost point*, and specific humidity.

Introduction

Water vapor varies significantly throughout the atmosphere. While a trace species in the middle atmosphere, it is the third most abundant gas in the Earth's lower troposphere. It has three important roles in weather and climate. First, water vapor is the dominant greenhouse gas so it has a significant effect on the planetary energy balance. Second, because the atmospheric capacity for water vapor varies exponentially with temperature, its radiative effects may act to amplify any surface warming; water vapor feedbacks are believed to roughly double anthropogenic warming (Intergovernmental Panel on Climate Change, 2007). Third, clouds and precipitation both begin as water vapor. Thus, water vapor acts indirectly on radiatively important clouds, while its condensation releases latent heat and affects precipitation. A major challenge in [weather forecasting](#) is improved precipitation forecasts, and water vapor is the atmospheric source for precipitation. Latent heat release represents about half the warming of the tropical atmosphere and makes a small but important contribution at higher latitudes, especially in storm systems (Hartmann, 1994). The importance of water vapor has made its observation a cornerstone of remote sounding techniques for decades.

Influence on weather

Thunderstorms and severe weather are significantly enhanced by the presence of water vapor, especially at low levels. Latent heat released by water vapor condensing onto cloud liquid droplets causes warming and reduced density, enhancing convective instability (Emanuel, 1994). This instability is most pronounced where cooler, dry air overlies warm, moist air, as is common in spring, summer, and fall in the American Midwest; high convective instability is a major factor in the formation of tornadoes. Latent heat release is also a major factor in organized tropical convective systems, including hurricanes.

Radiative effects and climate feedbacks

Water vapor is the dominant greenhouse gas in the atmosphere, with strongest absorption in the middle and upper

troposphere (400–100 hPa pressure) where many of its infrared spectral lines become saturated (Liou, 1992). Water vapor is not well mixed (unlike other greenhouse gases such as carbon dioxide or methane), so estimating its radiative effects has required direct observations of its distribution. This has presented a challenge, because only recently have high information content water vapor data sets become available from satellites (see below). The global upper tropospheric water vapor record prior to 2002 was based on regression models of local relative humidity versus satellite-observed brightness temperatures in the 6.3 μm infrared band from broadband radiometers (Soden et al., 2005). Other early studies addressed upper tropospheric water vapor variability using balloon-borne sensors, but only in the early twenty-first century did those sensors become sensitive enough to detect the very small water vapor amounts typical of the upper troposphere (Voemel et al., 2007).

Water vapor is also an important factor in climate feedbacks. Atmospheric water vapor has an enormous source at the ocean surface, depends strongly on temperature through the Clausius-Clapeyron relation, and acts as a greenhouse gas. Combined, these factors give water vapor a positive feedback (amplifying effect) on surface warming or cooling. Dessler et al. (2008) used satellite observations and El Niño-Southern Oscillation as a proxy for carbon dioxide-induced surface warming and examined radiative forcing kernel functions from climate models. They showed that climate models' water vapor feedback in response to warming roughly doubled surface warming, consistent with the observed atmospheric response (see also Dessler and Sherwood, 2009). In further confirmation of water vapor response to surface warming, Santer et al. (2007) attributed an increase in a 20 year record of total water vapor as a response to anthropogenic warming.

Direct confirmation of upper tropospheric water vapor response to surface warming – and verification of a positive water vapor feedback – remains a challenge (Boers and van Meijgaard, 2009; National Research Council, 2003). Also, the mechanisms whereby water vapor is mixed throughout the troposphere are not fully understood (Gambacorta et al., 2008). The vertical distribution of water vapor is important in feedback processes because lower tropospheric water vapor strongly couples surface changes to the atmosphere. Climate models are based on deep convective parameterizations, which can lead to biased water vapor distributions in models (Pierce et al., 2006) along with height-dependent temperature biases (John and Soden, 2007). Because water vapor and temperature are coupled by deep convection, these biases are related and their radiative contributions partly cancel (National Research Council, 2003; Bony et al., 2006). John and Soden (2007), following Held and Soden (2006), show evidence that climate model feedbacks are robust despite biases in mean fields. While the water vapor

feedback is well understood, parameterization of cloud physics in climate models (including the coupling between clouds and water vapor) is a major source of uncertainty in climate projection (Stephens, 2005; National Research Council, 2003).

Remote sensing of water vapor

The basis of all remote sensing or remote sounding systems is an instrument, or set of instruments, to observe electromagnetic radiation. (Remote sounding is the process of inferring vertical structure of temperature, clouds, and constituent gases like water vapor from observed spectral information.) In the case of water vapor, the observations are typically in the 6.3 μm infrared band or the 183 GHz microwave band. In addition to an instrument making observations, remote sensing systems usually include a numerical post-processing system to retrieve vertical distributions of geophysical parameters from the calibrated observed radiances. Therefore, remote sensing “observations” of water vapor (and many other atmospheric quantities) are in fact inferences about the state of the atmosphere. This makes the attribution of uncertainties an especially important and challenging component of remote sounding.

History of satellite remote sensing of water vapor

Satellite remote sensing of atmospheric water vapor extends back to the 1960s, when a variety of remote sensing instruments were launched into polar and geosynchronous orbits (Kidder and Vonder Haar, 1995). Many of these instruments included observations at wavelengths sensitive to water vapor. By the 1970s, basic methodologies for observing water vapor from space were established, and the United States began launching the TIROS Operational Vertical Sounder (TOVS) series of satellite instruments dedicated to tropospheric sounding. TOVS was first launched in 1978 and included the High-Resolution Infrared Sounder (HIRS) and the Microwave Sounding Unit. The highest-quality long-term satellite water vapor record is from Special Sensor Microwave Imager (SSM/I) instrument, though only as total water vapor. Other improvements in water vapor include the incorporation of four microwave water vapor channels with the Advanced Microwave Sounding Unit-B (AMSU-B) in 2000. AMSU-A, part of the Advanced TOVS sounders, included only a single water vapor channel. Technological advances have produced infrared instruments with higher spectral resolution than their microwave counterparts. This improved spectral resolution leads to greater information about water vapor vertical structure. The strength of microwave instruments is their greater coverage in the presence of clouds, though at the cost of vertical resolution. The Atmospheric Infrared Sounder (AIRS) instrument, launched in 2002 by the United States National Aeronautics and Space Administration (NASA) on the Aqua spacecraft into a 1:30 equator

crossing orbit, provides detailed information about water vapor (and a number of other quantities) from the surface to the upper troposphere with its 2000+ infrared channels and 20 microwave channels in a co-boresited AMSU-A instrument (Chahine et al., 2006). A similar instrument, the Infrared Atmospheric Sounding Interferometer (IASI), launched in 2006 by the European Space Agency provides similar information in a 9:30 local orbit on the first of three Meteorological Operational satellites (Centre National D'Etude Spatiales, 2010). IASI observations are collocated with microwave observations from the Microwave Humidity Sounder (Polar Orbiting Environmental Satellite, 2010) for additional constraints on the water vapor distribution. Other modern satellite instruments observing water vapor include the Microwave Limb Sounder (MLS), a limb-viewing instrument with sensitivity from the upper troposphere to the stratosphere and carried on the NASA Aura platform. This is a follow-on to an MLS instrument on the Upper Atmosphere Research Satellite (UARS) but with sensitivity to both temperature and water vapor (UARS MLS could detect variations in relative humidity only; see Fueglistaler for a review of instruments sensing the upper troposphere and stratosphere). Aura carries the Tropospheric Emission Spectrometer, with the unique ability to sense the water vapor isotopologue HDO in the troposphere (Worden et al., 2007).

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

- American Meteorological Society, 2010. Online glossary. <http://amsglossary.allenpress.com/glossary/>.
- Boers, R., and van Meijgaard, E., 2009. What are the demands on an observational program to detect trends in upper tropospheric water vapor anticipated in the 21st century? *Geophysical Research Letters*, doi:10.1029/2009GL040044.
- Bony, S., Colman, R., Kattsov, V. M., Allan, R. P., Bretherton, C. S., Dufresne, J.-L., Hall, A., Hallegatte, S., Holland, M. M., Ingram, W., Randall, D. A., Soden, B. J., Tselioudis, G., and Webb, M. J., 2006. How well do we understand and evaluate climate change feedback processes? *Journal of Climate*, **19**, 3445–3482.
- Centre National D'Etude Spatiales, 2010. <http://smc.cnes.fr/IASI/index.htm>.
- Chahine, M. T., Pagano, T. S., Aumann, H. H., Atlas, R., Barnett, C., Blaisdell, J., Chen, L., Divakarla, M., Fetzer, E. J., Goldberg, M., Gautier, C., Granger, S., Hannon, S., Irion, F. W., Kakar, R., Kalnay, E., Lambrigtsen, B. H., Lee, S.-Y., Le Marshall, J., McMillan, W. W., McMillin, L., Olsen, E. T., Revercomb, H., Rosenkranz, P., Smith, W. L., Staelin, D., Strow, L. L., Susskind, J., Tobin, D., Wolf, W., and Zhou, L., 2006. The atmospheric infrared sounder (AIRS): improving weather forecasting and providing new data on greenhouse gases. *Bulletin of the American Meteorological Society*, **87**, 911–926.
- Dessler, A. E., and Sherwood, S. C., 2009. A matter of humidity. *Science*, **323**, 1020–1021.
- Dessler, A. E., Zhang, Z., and Yang, P., 2008. Water-vapor climate feedback inferred from climate fluctuations, 2003–2008. *Geophysical Research Letters*, doi:10.1029/2008GL035333.
- Emanuel, K., 1994. *Atmospheric Convection*. New York: Oxford University Press. Available online: <http://books.google.com/books>.
- Fueglistaler, S., Dessler, A. E., Dunkerton, T. J., Folkins, I., Fu, Q., and Mote, P. W., 2009. Tropical tropopause layer. *Reviews of Geophysics*, doi:10.1029/2008RG000267.
- Gambacorta, A., Barnett, C., Soden, B., and Strow, L., 2008. An assessment of the tropical humidity-temperature covariance using AIRS. *Geophysical Research Letters*, doi:10.1029/2008GL033805.
- Hartmann, D. L., 1994. *Global Physical Climatology*. San Diego: Academic.
- Held, I. M., and Soden, B. J., 2006. Robust responses of the hydrological cycle to global warming. *Journal of Climate*, **19**, 5686–5699.
- Intergovernmental Panel on Climate Change, 2007. Summary for policymakers. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Avery, K. B., Tignor, M., and Miller, H. L. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, NY: Cambridge University Press.
- John, V. O., and Soden, B. J., 2007. Temperature and humidity biases in global climate models and their impact on climate feedbacks. *Geophysical Research Letters*, doi:10.1029/2007GL030429.
- Kidder, S. Q., and Vonder Haar, T. H., 1995. *Satellite Meteorology: An Introduction*. San Diego: Academic.
- Liou, K. N., 1992. *Radiation and Cloud Processes in the Atmosphere*. New York: Oxford University Press.
- National Research Council, 2003. *Understanding Climate Change Feedbacks*. Washington, DC: National Academies.
- Pierce, D. W., Barnett, T. P., Fetzer, E. J., and Gleckler, P. J., 2006. Three-dimensional tropospheric water vapor in coupled climate models compared with observations from the AIRS satellite system. *Geophysical Research Letters*, doi:10.1029/2006GL027060.
- Polar Orbiting Environmental Satellite, 2010. <http://goespoes.gsfc.nasa.gov/poes/instruments/mhs.html>.
- Santer, B. D., Mears, C., Wentz, F. J., Taylor, K. E., Gleckler, P. J., Wigley, T. M. L., Barnett, T. P., Boyle, J. S., Bruggemann, W., Gillett, N. P., Klein, S. A., Meehl, G. A., Nozawa, T., Pierce, D. W., Stott, P. A., Washington, W. M., and Wehner, M. F., 2007. Identification of human-induced changes in atmospheric moisture content. *Proceedings of the National Academy of Sciences*, **104**(39), 15248–15253.
- Soden, B. J., Jackson, D., Ramaswamy, V., Schwarzkopf, M. D., and Huang, X., 2005. The radiative signature of upper tropospheric moistening. *Science*, **310**, 841–844.
- Stephens, G. L., 2005. Cloud feedbacks in the climate system: a critical review. *Journal of Climate*, **18**, 237–273.
- Voemel, H., David, D. E., and Smith, K., 2007. Accuracy of tropospheric and stratospheric water vapor measurements by the cryogenic frost point hygrometer: instrumental details and observations. *Journal of Geophysical Research*, doi:10.1029/2006JD007224.
- Wallace, J. M., and Hobbs, P. V., 2006. *Atmospheric Science: An Introductory Survey*. San Diego: Academic.
- Worden, J., Noone, D., and Bowman, K., 2007. Importance of rain evaporation and continental convection in the tropical water cycle. *Nature*, **445**, 528–532.

WEATHER PREDICTION

Peter Bauer

European Centre for Medium-Range Weather Forecasts
(ECMWF), Reading, UK

Synonyms

Weather forecasting

Definition

Weather. State of the atmosphere and its day-to-day variation, mostly described by temperature, wind, cloudiness, and precipitation.

Weather Prediction. Prediction of weather at a given time and location using numerical models and observations.

Numerical atmospheric modeling

Numerical models are used to simulate the evolution of all those processes in the atmosphere and at surfaces that affect the atmospheric state. Increasingly, these models add complexity to the modeled components of atmosphere and surfaces that go beyond the basic formulation of mass, heat, and momentum transport. With increasing complexity of the models, a wider range of spatial and temporal process scales has to be accounted for and the diversity and nonlinearity of the modeled processes increases as well (Kalnay, 2003).

In many applications, numerical models or model output are combined, for example, by nesting smaller-scale models into larger-scale models; by coupling of models for the atmosphere and land surfaces (see “*Ocean-Atmosphere Water Flux and Evaporation*”; “*Land-Atmosphere Interactions, Evapotranspiration*”), hydrology, waves, and oceans; and by coupling of atmospheric dynamics with chemistry models.

Global numerical models usually employ a set of primitive equations to describe atmospheric dynamics under the assumption of momentum conservation, heat exchange through thermodynamics, and mass conservation (see “*Atmospheric General Circulation Models*”). Processes with scales well below the scales that the numerical model can resolve are described by physical parameterizations that approximate the effect of sub-grid processes on the resolved scales and vice versa. Among these are radiation processes (see “*Radiation, Electromagnetic*”), cloud condensation and convection, orographic drag, turbulent diffusion, and most surface processes (Kalnay, 2003).

Since most processes exhibit nonlinear behavior and are discretized in space and time, the corresponding equations are solved numerically. The choice of solution method depends on the model type. Some global models are spectral models, where the horizontal dimension is described by a set of waves and where the horizontal resolution is proportional to the highest defined wave number. Others solve the above equations on structured or unstructured horizontal grids. The vertical dimension is

usually defined by finite layers with varying choice of coordinate systems (e.g., sigma, eta, theta, hybrid coordinates; e.g., Warner, 2011).

At operational numerical weather prediction (NWP) centers, numerical models are run as part of the analysis system that estimates the state of the (global) atmosphere at a given time as well as for producing the forecast over the desired time range. The analyses are used to initialize the forecast model runs. These deterministic forecasts are often complemented by ensemble forecasts for which a set of forecasts are produced from perturbed initial conditions, possibly using further stochastic input. The perturbations are supposed to represent the uncertainty of the initial state estimate as well as the model, so that the ensemble of forecasts provides an estimate of forecast uncertainty. To reduce computational cost, the ensemble forecasts are run at lower spatial resolutions than the deterministic model. With increasing forecast range (monthly/seasonal), ensemble-type modeling becomes more important because the nonlinear response to small differences in the initial conditions and model errors become large and the knowledge of ensemble forecast spread is crucial for interpreting forecast skill.

Atmospheric data assimilation

Data assimilation systems (see “*Data Assimilation*”) in NWP provide the mathematical framework for performing atmospheric analyses that represent the best estimate of the state of the atmosphere at a certain time. The quality of the forecast depends on both the accuracy of the numerical model and the accuracy of the initial state estimate. The initial state estimate is called the analysis and represents an inversion problem. In general, this inversion problem is underdetermined so that the analysis must employ information from a priori data (in NWP usually short-range forecasts initialized with previous analyses) and observations using a mathematical framework to optimally combine the two.

The complexity of the data assimilation system to be used depends on the application and the affordable computational cost and can range from simple interpolation techniques to four-dimensional variational and ensemble Kalman filter schemes or even nonlinear methods (Daley, 1991; Ide et al., 1997; Rabier, 2005). Global analysis systems are solving the above inversion problem with state vector dimensions of the order of 10^8 (the product of the number of grid points, number of levels, and dimension of state vector) and observation vector dimensions of the order of 10^7 (product of number of observation points, number of levels, or satellite instrument channels/range gates).

Today, most global operational NWP centers are operating 4D-Var (Lewis and Derber, 1985) data assimilation systems that are capable of combining sparse and heterogeneously distributed data with a dynamical model. A computationally efficient derivative is the incremental 4D-Var method that is based on the assumption that model behavior is nearly linear in the vicinity of a good

short-range forecast of the model state (first guess) so that the analysis is produced from incrementally updating the first-guess estimate (Bouttier and Courtier, 2002; Rawlins et al., 2007). The advantage of 4D-Var methods lies in the fact that (1) they produce meteorological fields that are dynamically consistent because the optimization is performed over a time window through which the model is integrated and (2) that computationally efficient adjoint models can be used with incremental methods (Courtier et al., 1993). Systems like this are currently in use at ECMWF, the UK Met Office, Météo-France, the Meteorological Service of Canada (MSC), the Japan Meteorological Agency (JMA) (for the regional model), and in the USA at the Naval Research Laboratory (NRL).

At regional scales, ensemble-based methods (Andersen, 2001; Houtekamer and Mitchell, 2001) become increasingly implemented because they do not require the development of adjoint models and are capable of explicitly calculating analysis error statistics than can be used for estimating model errors and for initializing ensembles of forecasts (Lorenc, 2003). Due to the necessity of running ensembles, the computational effort is considerable and currently not affordable at global scales with the same horizontal resolutions as obtained with variational assimilation methods.

However, increasingly hybrid systems composed of high-resolution variational and low-resolution ensemble data assimilation are implemented in which the ensemble system provides background error statistics for the high-resolution analysis.

Satellite data in weather prediction

Satellite data products

Satellite data can be assimilated as level 1 (e.g., calibrated and geo-located electromagnetic radiances) or level 2 (e.g., derived geophysical products such as temperature and humidity; see “*Geophysical Retrieval, Overview*”) data. The choice depends on various factors, most prominently on the amount of maintenance required in an operational system. Level 2 product retrieval often employs a similar inversion framework as NWP data assimilation system to derive the desired parameters, namely, a priori information from NWP models or climatologies, radiative transfer, error, and bias models. Most level 2 products, however, employ different models and a priori constraints than used in NWP modeling, and their error characteristics are often not well defined or difficult to account for in data assimilations systems (Joiner and Dee, 2000). Further, when retrieval algorithms or instrument characteristics (e.g., loss of channels, increase of noise, and instrument recalibration) change, the NWP data assimilation system must be tuned to the performance of the new product, which is often cumbersome in an operational framework.

On the other hand, level 1 satellite data assimilation requires running radiative transfer models (see “*Radiative Transfer, Solution Techniques*”; “*Radiative Transfer, Theory*”) that simulate observation-equivalent radiances

from the model state. For the most important application of satellite data in NWP, that is, the observation of temperature and moisture structures in the atmosphere, radiative transfer models are very fast and accurate (Saunders et al., 1999; Han et al., 2006), and they allow the flexible use of radiometer channels as a function of situation-dependent sensitivity and potential channel corruption. They also greatly simplify observation error and bias estimation.

The same approach of level 1 data assimilation is increasingly used for observations sensitive to clouds, precipitation, aerosols, surface characteristics and trace gases and observables from active sensors.

History of satellite data usage

Derived wind vectors obtained from geostationary cloud (later also water vapor) feature tracking were among the first products used in the early days of satellite data assimilation. The associated cloud feature heights are retrieved from infrared window channels (Nieman et al., 1993), and the data produced good impact on analysis and forecast quality, mainly in the Southern Hemisphere where little conventional observations were available. With the launch of infrared and microwave sounders onboard polar orbiting satellites, more information on vertical temperature structures became available. At that time, the assimilation of retrieved geophysical products was preferred over radiances due to less efficient radiative transfer models, more simple observation operators, and the uncertain impact of satellite data in general. Initial experiments with sounder data produced even negative impact due to unknown bias characteristics of the retrieved profiles and model error statistics that were not tuned to deal with profile data obtained from observations other than radiosondes.

An intermediate step between the assimilation of retrieved products and radiance assimilation is based on 1D-Var techniques. Here, a single-column (one-dimensional variational or 1D-Var) retrieval is performed with radiance observations, and the retrieval product is assimilated globally by the corresponding global analysis scheme (optimum interpolation, 3D/4D-Var). The advantage of such an approach is that the constraints and assumptions used in the 1D-Var retrieval are very similar to those used in the global scheme because they are performed within the same analysis system. One of the earliest protagonists of this approach were Eyre et al. (1993) in Europe, producing retrieved temperature profiles from TOVS data that comprises observations from the HIRS, MSU, and SSU instruments. This approach has greatly facilitated the promotion of satellite data usage in NWP and was later extended to the use of moisture-sensitive channels and instruments such as the AMSU-B and the SSM/I (Phalippou, 1996). Most of these systems were later replaced by direct radiance assimilation, that is, excluding the intermediate retrieval step, for the abovementioned reasons (Andersson et al., 1994; Derber and Wu, 1998).

The initial concerns over general satellite data impact were mostly overcome by the time 4D-Var data assimilation systems were established, mainly because of the improved treatment of spatial and temporal collocation between data and model simulations, better forecast model error formulations, and the improved interaction of temperature and moisture with model dynamics (Andersson and Thépaut, 2008). In the 1990s and early years of the twenty-first century, the dominant impact of satellite data on numerical weather forecast skill was contributed by ATOVS data that combines HIRS AMSU-A and AMSU-B observations. The first ATOVS sensor package was launched with NOAA-15 in 1998 and has been continued until NOAA-19 (launched in 2009). For both NOAA-18 and NOAA-19, the AMSU-B has been replaced with MHS. Since 2006, the same package is also available on the EPS series METOP.

The microwave sounder system is complemented by, so-called, microwave imagers (e.g., SSM/I, and the follow-on instrument SSMIS, AMSR-E onboard Aqua, TMI onboard TRMM, AMSR-2 onboard GCOM-W) that contribute information on sea-surface temperature, near-surface wind speed, integrated atmospheric moisture, clouds, and precipitation (see “*Microwave Radiometers*”).

A major step forward has been the development of infrared spectrometers (grating spectrometer AIRS onboard EOS Aqua in 2002, interferometer IASI onboard METOP-A/B in 2006/2012 and CrIS onboard Suomi NPP in 2011) that provide unprecedented vertical resolution and measurement accuracy by making available thousands of channels with very high spectral resolution covering the 4–15 μm range. These instruments have produced substantial impact on NWP (Collard and McNally, 2009) and will complement microwave sounding radiometers for the next 20 years. Due to their superior spectral resolution, these instruments are also crucial for atmospheric chemistry and air quality applications (Clerbaux et al., 2009).

Further, the utilization of information on atmospheric temperature and moisture structures obtained from the bending of rays of actively transmitted radio waves by the GNSS has vastly increased over the past 5 years (Eyre, 1994; Healy and Thépaut, 2006; see “*Limb Sounding, Atmospheric*”; “*GPS, Occultation Systems*”). Currently, the assimilation of GPS signals in occultation mode by dedicated receivers (onboard CHAMP, GRAS onboard METOP, COSMIC Formosat-3 constellation) provides radiosonde-equivalent measurement accuracy, mainly in the upper atmosphere where NWP models tend to exhibit significant errors.

Apart from temperature structures in the upper atmosphere, trace gas observations (mainly ozone) can provide valuable contributions to weather prediction since ozone has a strong impact on radiative heating (see “*Stratospheric Ozone*”). Solar backscattered ultraviolet instruments (e.g., SBUV, OMI, and GOME-1/2) provide

the bulk of the observations when sunlight is available, while infrared spectrometers complement the observations at nighttime.

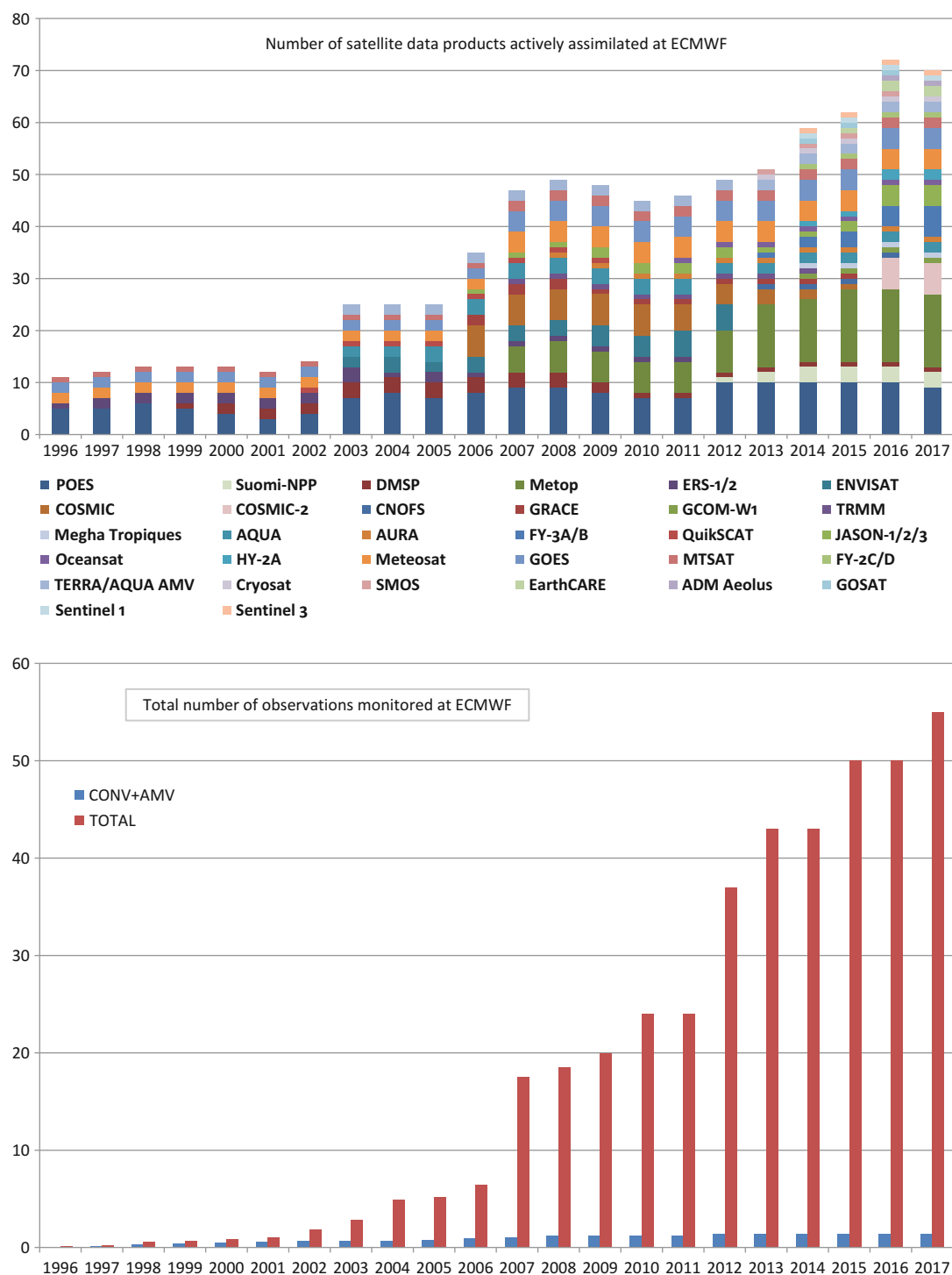
Apart from clear-sky satellite data, efforts toward cloud-affected data assimilation have been successful in recent years (Bauer et al. 2010; Geer et al. 2010; Bauer et al. 2011. Bauer et al., 2006a, b). This was mainly achieved by the greatly improved global model moist physical parameterizations and the enhanced computational capabilities that allow the operational employment of multiple scattering radiative transfer models (Greenwald et al., 2002; Bauer et al., 2006c). The explicit treatment of clouds and precipitation in operational analysis systems is accompanied by a large set of uncertainties, for example, greater model nonlinearity, potential dynamic instabilities, large and unknown error structures, as well as unknown model biases (Errico et al., 2007).

Atmospheric modeling requires accurate constraints of energy and water fluxes at the interface with land and ocean surfaces. Over oceans, the interaction between wind and waves is treated by wave models. Here, near-surface wind observations from passive (microwave radiometers) or active (scatterometers; see “*Radar, Scatterometers*”) satellite data and direct observations of wave height from altimeters (see “*Radar, Altimeters*”) and directional wave spectra from synthetic aperture radars (see “*Radar, Synthetic Aperture*”) are part of the operational set of assimilated data.

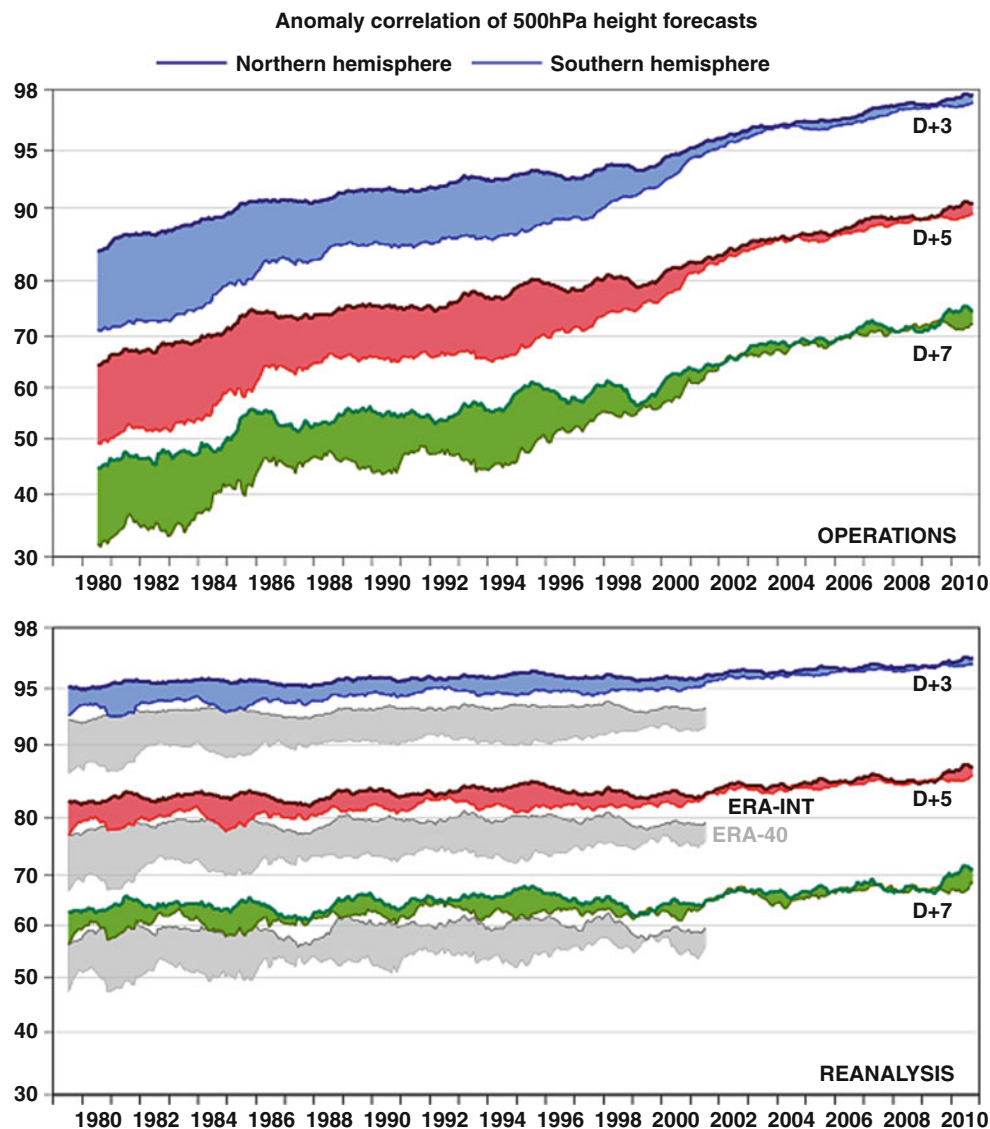
Over land surfaces, a recent development in NWP is the use of microwave radiometer observations to constrain soil moisture analysis (see “*Soil Moisture*”). This can be accomplished by fixed aperture radiometer observations at 6–10 GHz from TMI and AMSR-E (Reichle et al., 2007) and moderate resolution observations at 1.4 GHz from synthetic aperture imagery by SMOS (Merlin et al., 2006). Note that over land surfaces, RFI can impose serious limitations on data quality and usefulness for NWP (see “*Radio-Frequency Interference (RFI) in Passive Microwave Sensing*”). Also scatterometer is used to extract information on soil properties.

Snow and sea-ice products, mostly obtained from passive and active microwave instruments, provide important information for NWP systems on how to define surface albedo and heat fluxes.

Figure 1 illustrates the dramatic increase of satellite data diversity and volume since 1996 as well as a prediction until 2017 at ECMWF. The historic evolution of used data does not necessarily reflect the sequential launch of individual satellites but rather the ability of NWP data assimilation systems and computers to digest the available information. Today, data from about 50 different instruments is used constraining geophysical parameters in the atmosphere and at both land and ocean surfaces. The gap between data volume contributed by conventional (i.e., nonsatellite) and satellite data has largely widened. More



Weather Prediction, Figure 1 History and prediction of satellite data usage in data assimilation at ECMWF in terms of instruments (*top*) and data volume per day for conventional and satellite observations (*bottom*).



Weather Prediction, Figure 2 Evolution of ECMWF 500 hPa height forecast skill score, expressed as anomaly correlation. *Top panel* shows day 3 (blue), day 5 (red), day 7 (green) scores between Northern (thick lines) and Southern (thin lines) Hemispheres. *Bottom panel* shows corresponding scores from ECMWF reanalysis, namely, ERA-40 (gray) and ERA-Interim (colors).

than 40 million satellite observations are monitored per day, 95 % of which are assimilated directly as radiances while the remainder is composed of retrieved products.

Impact of satellite data on prediction skill

Data impact in NWP systems can be quantified in various ways and by assessing the impact on both analyses and forecasts. It is assumed that better analyses will provide better initial conditions for forecasts. Analysis quality is usually quantified by the comparison of NWP model fields to all available observations before and after they have been assimilated. A better system is expected to produce a consistently better fit of the analysis to the

observations and should maintain this advantage up to the next short-range forecast that has been initialized with this analysis. Forecast skill can be quantified by comparing forecasts to both observations and analyses.

Figure 2 shows the evolution of the ECMWF forecast model skill over the period 1980–2010 for the 3, 5, and 7 day forecasts over the Northern and Southern Hemispheres from the operational system (Figure 2a) and from the ECMWF reanalyses ERA-40 (gray, Uppala et al., 2005) and ERA-Interim (in colors, Uppala et al., 2008). Figure 2a illustrates the substantial increase of skill over three decades and the strong reduction of the difference between skill over the Northern and Southern

Hemispheres. While the former is the result of the combined evolution of model physics, data assimilation, and observing systems, the latter is only explained by the contribution of satellite data due to the sparseness of conventional data over Southern Hemispheric oceans. [Figure 2b](#) provides another angle at disentangling the individual contributions to the time series of NWP forecast skill: The reanalyses have been produced with a model version that has been frozen in 2001 (ERA-40) and 2006 (ERA-Interim), respectively. The small increase of skill over the ERA period is mostly due to the advancement of satellite observations, while the difference between ERA-40 and ERA-Interim is mainly due to the improvement of model and data assimilation system between 2001 and 2006. The larger gap of Northern and Southern Hemisphere forecast skill of ERA-40 compared to ERA-Interim in the overlap period between 1989 and 2001 illustrates the fact that the data assimilation system of ERA-Interim exploits the same satellite observations more effectively than ERA-40.

If the impact of individual observing systems (satellites or instruments) is evaluated, the most prominent assessment tool is the so-called Observing System Experiment (OSE), in which new data is added to an existing system and the relative difference to a control system is evaluated. Similarly, individual data sources can be withdrawn from a full system to assess the impact of the withdrawn data.

More sophisticated methods involve the model operators that are used in the data assimilation system. Based on forecast error estimates from the difference between forecasts and verifying analysis, the model and observation operator adjoints employed in 4D-Var can be used to deduce the dependence of this forecast error on individual observation types that were used in the initializing analysis (Zhu and Gelaro, 2008). An alternative is the use of ensemble-based analysis and forecasting techniques that evaluate forecast impact from ensemble spread (that represents model error) with or without specific observation types. Finally, Observing System Simulation Experiments (OSSE) provide a framework for evaluating the potential impact of observations that do not yet exist and therefore require an observation simulation from independent NWP models (Arnold and Dey, 1986; Tan et al., 2007).

A major OSE impact study was conducted in 2006–2007 to evaluate the impact of the satellite observing system in global NWP at ECMWF (Kelly and Thépaut, 2007). The experiments demonstrated that (a) infrared spectrometers (AIRS, IASI) produce the largest impact per single instrument on geopotential height and temperature forecast skill and (b) that the currently available constellation of 4–5 microwave sounders (AMSU-A/B/MHS) produces a very similar relative impact compared to one advanced infrared sounder. The results are similar for the Northern Hemisphere but with a smaller dynamic range due to the stronger constraint from denser conventional observations obtained over the continents. Further studies

suggest that, apart from infrared and microwave sounders, GPS radio-occultation and scatterometer data produce significant contributions to forecast accuracy since they are most directly related to temperature and surface wind (pressure) with good global coverage.

In terms of atmospheric moisture, Kelly and Thépaut (2007) confirmed previous investigations showing that SSM/I data has the strongest impact in the lower troposphere over oceans complemented by AMSU-B data in the mid and upper troposphere (Andersson et al., 2007). The impact of clear-sky microwave imager data is about as strong as that of cloud-affected data (Kelly et al., 2008).

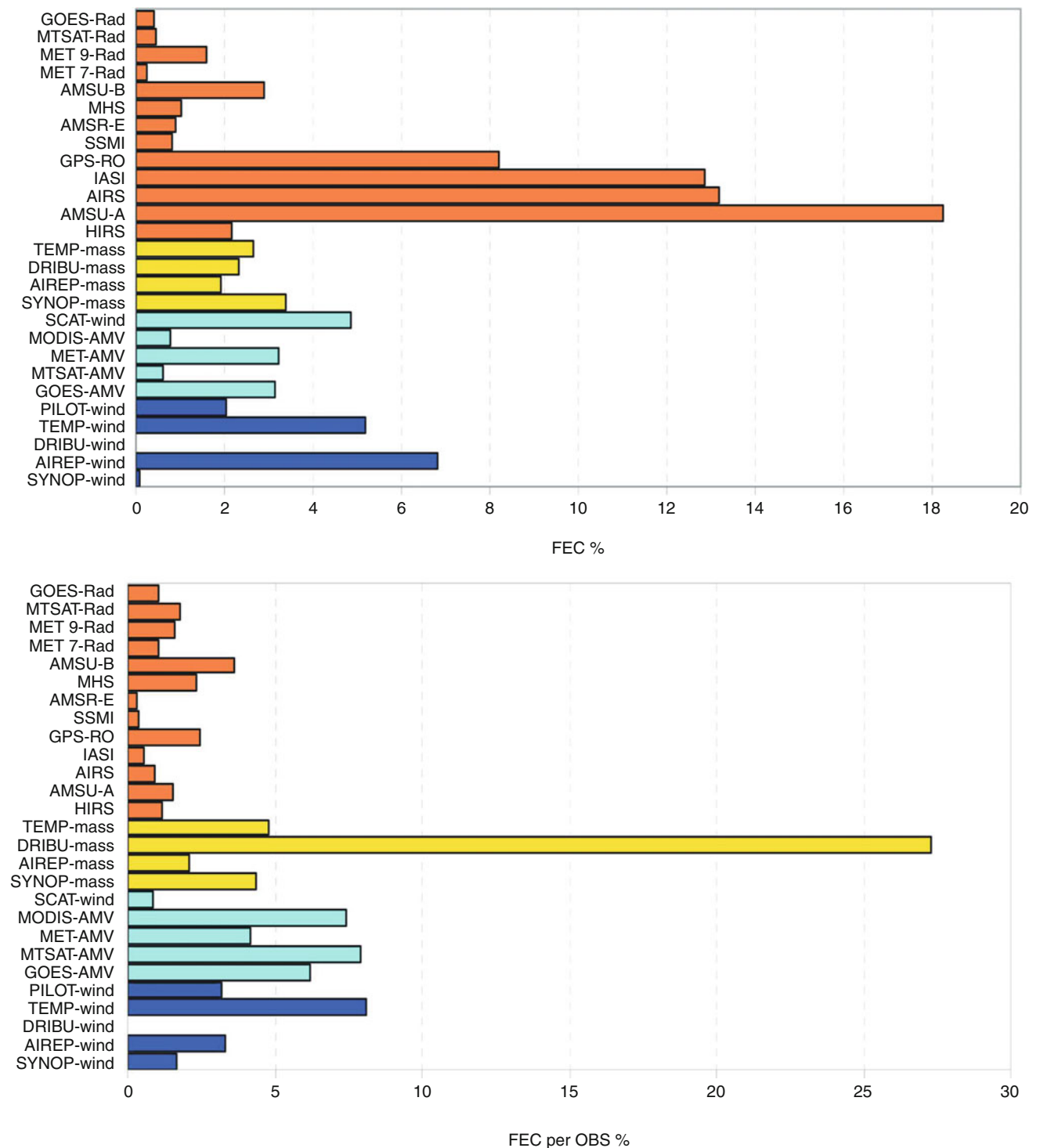
[Figure 3](#) shows a different measure of global forecast impact from selected observation types that is obtained from the 24 h forecast error sensitivity to the accumulated sensitivity to all observation types (Cardinali, 2009). The methodology is able to estimate observational impact without having to add/withdraw them but only applies to short-range forecast impact estimation. In [Figure 3a](#), the total impact of the most prominent satellite and conventional observations for a 4 month period (September–December 2008) is shown, while [Figure 3b](#) shows the impact per individual observation.

It is evident that infrared spectrometers and microwave sounders produce the strongest impact followed by radio-occultation observations. [Figure 3b](#) shows that surface observations of pressure over the oceans from drifting buoys but also all types of direct wind observations have a strong impact, which suggests the importance of accurate wind observations from satellites as expected from ADM/Aeolus in the future (see “[Lidar Systems](#)”).

It is important to note that the impact of individual observations depends on the NWP system and the weight assigned to the observations in the analysis. It is therefore crucial to evaluate the NWP model, the data assimilation system, and the entire set of used observations together to characterize the importance of existing satellite data for NWP and to estimate the potential impact of future systems.

Summary

Current weather forecast skill is strongly driven by the sophistication of the physical processes represented in numerical models and advanced data assimilation schemes allowing vast amounts of data from conventional sources and satellites to be used. Globally, more than 40 million observations per day are used from about 50 different satellite instruments to produce atmospheric analyses with which the forecast models are initialized. The most important instruments are passive radiometers that measure infrared and microwave radiation emitted by the surface-atmosphere system and that are mostly exploited to derive information on temperature and moisture structures. Increasingly, observations of clouds and rain, surface waves, land surface characteristics, and atmospheric trace gases are added. In parallel, numerical models become increasingly capable of representing more



Weather Prediction, Figure 3 Relative contribution of different observing systems to 24 h forecast error reduction for September–December 2008 (Cardinali, 2009). *Top panel* shows contribution per observing system; *bottom panel* shows contribution per single observation.

complex physical and chemical processes at smaller scales. Ensemble analysis and forecasting systems allow the estimation of analysis and forecasting uncertainties – a crucial information in forecasting highly nonlinear atmospheric phenomena. Future satellite observing systems will develop toward more hyper-spectral instruments covering wider spectral ranges with fine spectral resolution as well as active instruments that sample vertical structures and wind very accurately.

Abbreviations

4D-Var–Four-Dimensional Variational Assimilation
 AIRS–Atmospheric Infrared Sounder
 IASI–Infrared Atmospheric Sounding Interferometer
 ATOVS–Advanced TIROS Operational Vertical Sounder
 TOVS–TIROS Operational Vertical Sounder
 HIRS–High-Resolution Infrared Sounder
 AMSU-A–Advanced Microwave Sounding Unit A
 AMSU-B–Advanced Microwave Sounding Unit B
 MHS–Microwave Humidity Sounder
 SSM/I–Special Sensor Microwave/Imager
 SSMIS–Special Sensor Microwave Imager Sounder
 TMI–TRMM Microwave Imager
 TRMM–Tropical Rainfall Measuring Mission
 AMSR-E–Advanced Microwave Scanning Radiometer E
 METOP–Meteorological Operational Polar satellite
 EPS–EUMETSAT Polar System
 NOAA–National Oceanic and Atmospheric Administration
 GNSS–Global Navigation Satellite System
 GRAS–GNSS Receiver for Atmospheric Sounding
 CHAMP–Challenging Minisatellite Payload
 GPS–Global Positioning System
 NWP–Numerical Weather Prediction
 ERA–ECMWF Reanalysis
 ECMWF–European Center for Medium-Range Weather Forecasts
 SMOS–Soil Moisture Ocean Salinity
 EOS–Earth Observing System
 COSMIC–Constellation Observing System for Meteorology, Ionosphere, and Climate
 OSE–Observing System Experiment
 OSSE–Observing System Simulation Experiment
 Suomi NPP–Suomi National Polar-orbiting Partnership
 CrIS–Cross-track Infrared Sounder
 GCOM-W–Global Change Observation Mission - Water

Bibliography

- Andersen, J. L., 2001. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, **129**, 2884–2902.
- Andersson, E., and Thépaut, J.-N., 2008. ECMWF's 4D-Var data assimilation system – the genesis and ten years in operations. *ECMWF Newsletter*, **115**, 8–12.
- Andersson, E., Pailleux, J., Thépaut, J.-N., Eyre, J., McNally, A. P., Kelly, G. A., and Courtier, P., 1994. Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **120**, 627–653.
- Andersson, E., Hólm, E., Bauer, P., Beljaars, A., Kelly, G. A., McNally, A. P., Simmons, A. J., Thépaut, J.-N., and Tompkins, A. M., 2007. Analysis and forecast impact of the main humidity observing system. *Quarterly Journal of the Royal Meteorological Society*, **133**, 1473–1485.
- Arnold, C. P., and Dey, C. H., 1986. Observing-system simulation experiments: past, present and future. *Bulletin of the American Meteorological Society*, **67**, 687–695.
- Bauer, P., Auligne, T., Bell, W., Geer, A. J., Guidard, V., Heilliette, S., Kazumori, M., Kim, M.-J., McNally, A. P., Macpherson, B., and Renshaw, R., 2011. Status of satellite cloud and precipitation assimilation at operational NWP centres. *Quarterly Journal of the Royal Meteorological Society*, **137**, 1934–1951.
- Bauer, P., Geer, A., Lopez, P., and Salmond, D., 2010. Direct 4D-Var assimilation of all-sky radiances. Part I: Implementation. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1868–1885.
- Bauer, P., Lopez, P., Benedetti, A., Salmond, D., and Moreau, E., 2006a. Implementation of 1D + 4D-Var assimilation of precipitation affected microwave radiances at ECMWF, part I: 1D-Var. *Quarterly Journal of the Royal Meteorological Society*, **132**, 2277–2306.
- Bauer, P., Lopez, P., Benedetti, A., Salmond, D., Saarinen, S., and Bonazzola, M., 2006b. Implementation of 1D + 4D-Var assimilation of precipitation affected microwave radiances at ECMWF, part II: 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **132**, 2307–2332.
- Bauer, P., Moreau, E., Chevallier, F., and O'Keefe, U., 2006c. Multiple-scattering microwave radiative transfer for data assimilation applications. *Quarterly Journal of the Royal Meteorological Society*, **132**, 1259–1281.
- Bouttier, F., and Courtier, P., 2002. *Data Assimilation Concepts and Methods*. ECMWF Training course lectures, available from ECMWF, Shinfield Park, Reading, UK, 59 p.
- Cardinali, C., 2009. Monitoring the observation impact on the short-range forecast. *Quarterly Journal of the Royal Meteorological Society*, **135**, 239–250.
- Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F., 2009. Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder. *Atmospheric Chemistry and Physics*, **9**, 6041–6054.
- Collard, A., and Healy, S., 2003. The combined impact of future space-based atmospheric sounding instruments on numerical weather prediction analysis fields: a simulation study. *Quarterly Journal of the Royal Meteorological Society*, **129**, 2741–2760.
- Collard, A., and McNally, A. P., 2009. The assimilation of infrared atmospheric sounding interferometer radiances at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, **135**, 1044–1058.
- Courtier, P., Derber, J., Errico, R., Luis, J.-F., and Vukicevic, T., 1993. Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus*, **45A**, 342–357.
- Daley, R., 1991. *Atmospheric Data Analysis*. Cambridge: Cambridge University Press. Cambridge Atmospheric and Space Science Series. 457 pp. ISBN 0-521-38215-7.

- Derber, J. C., and Wu, W.-S., 1998. The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Monthly Weather Review*, **8**, 2287–2302.
- Errico, R., Bauer, P., and Mahfouf, J.-F., 2007. Assimilation of cloud and precipitation data: current issues and future prospects. *Journal of the Atmospheric Sciences*, **64**, 3789–3802.
- Eyre, J. R., 1994. *Assimilation of Radio Occultation Measurements into a Numerical Prediction System*. Reading, Berks: ECMWF Tech. Memo No. 199.
- Eyre, J., Kelly, G. A., McNally, A. P., Andersson, E., and Persson, A., 1993. Assimilation of TOVS radiance information through one-dimensional variational analysis. *Quarterly Journal of the Royal Meteorological Society*, **119**, 1427–1463.
- Geer, A., Bauer, P., and Lopez, P., 2010. Direct 4D-Var assimilation of all-sky radiances. Part II: assessment. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1886–1905.
- Greenwald, T. J., Hertenstein, R., and Vukicevic, T., 2002. An all-weather observational operator for radiance data assimilation with mesoscale forecast models. *Monthly Weather Review*, **130**, 1882–1897.
- Han, Y., van Delst, P., Liu, Q., Weng, F., Yan, B., Treadon, R., and Derber, J., 2006. *Community Radiative Transfer Model (CRTM) – Version 1*. NOAA/NESDIS Tech. Rep. 122 pp.
- Healy, S., and Thépaut, J.-N., 2006. Assimilation experiments with CHAMP GPS radio occultation measurements. *Quarterly Journal of the Royal Meteorological Society*, **132**, 605–623.
- Houtekamer, P. L., and Mitchell, H. L., 2001. A sequential ensemble kalman filter for atmospheric data assimilation. *Monthly Weather Review*, **129**, 123–137.
- Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C., 1997. Unified notation for data assimilation: operational, sequential and variational. *Journal of Meteorological Society of Japan*, **75**, 181–189.
- Joiner, J., and Dee, D. P., 2000. An error analysis of radiance and suboptimal retrieval assimilation. *Quarterly Journal of the Royal Meteorological Society*, **126**, 1495–1514.
- Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge: Cambridge University Press.
- Kelly, G. A., and Thépaut, J.-N., 2007. Evaluation of the impact of the space component of the global observing system through observing system experiments. *ECMWF Newsletter*, **113**, 16–28.
- Kelly, G. A., Bauer, P., Geer, A., Lopez, P., and Thépaut, J.-N., 2008. Impact of SSM/I observations related to moisture, clouds and precipitation on global NWP forecast skill. *Monthly Weather Review*, **136**, 2713–2726.
- Kursinski, E., Hajj, G., Bertiger, W., Leroy, S., Meehan, T., Romans, L., Schofield, J., McCleese, D., Melbourne, W., Thornton, C., Yunck, T., Eyre, J., and Nagatani, R., 1996. Initial results of radio occultation observations of earth's atmosphere using the global positioning system. *Science*, **271**, 1107–1110.
- Kursinski, E., Hajj, G., Schofield, J., Linfield, R., and Hardy, K., 1997. Observing earth's atmosphere with radio occultation measurements using the global positioning system. *Journal of Geophysical Research*, **102**, 23,429–23,465.
- Lewis, J. M., and Derber, J. C., 1985. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37**, 309–327.
- Lorenc, A. C., 2003. The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **129**, 3183–3203.
- Luntama, J.-P., Kirchengast, G., Borsche, M., Foelsche, U., Steiner, S., Healy, S., von Engeln, A., O'Clerigh, E., and Marquardt, C., 2008. Prospects of the EPS GRAS mission for operational atmospheric applications. *Bulletin of the American Meteorological Society*, **89**, 1863–1875.
- Matricardi, M., Chevallier, F., Kelly, G. A., and Thépaut, J.-N., 2004. An improved general fast radiative transfer model for the assimilation of radiance observations. *Quarterly Journal of the Royal Meteorological Society*, **130**, 153–173.
- Merlin, O., Chehbouni, A., Boulet, G., and Kerr, Y., 2006. Assimilation of disaggregated microwave soil moisture into a hydrologic model using coarse-scale meteorological data. *Journal of Hydrometeorology*, **7**, 1308–1322.
- Nieman, S., Schmetz, J., and Menzel, P., 1993. A comparison of several techniques to assign heights to cloud tracers. *Journal of Applied Meteorology*, **32**, 1559–1568.
- Phalippou, L., 1996. Variational retrieval of humidity profile, wind speed and cloud liquid water path with the SSM/I: potential for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, **122**, 327–355.
- Poli, P., Moll, P., Puech, D., Rabier, F., and Healy, S., 2009. Quality control, error analysis, and impact assessment of FORMOSAT-3/COSMIC in numerical weather prediction. *Terrestrial Atmospheric and Oceanic Sciences*, **20**, 101–113.
- Rabier, F., 2005. Overview of global data assimilation developments in numerical weather prediction centres. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3215–3233.
- Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D., Inverarity, G. W., Lorenc, A. C., and Payne, T. J., 2007. The Met office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, **133**, 347–362.
- Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P. P., Njoku, E. G., and Owe, M., 2007. Comparison and assimilation of global soil moisture retrievals from the advanced microwave scanning radiometer for the earth observing system (AMSR-E) and the scanning multichannel microwave radiometer (SMMR). *Journal of Geophysical Research*, **112**, D09108, doi:10.1029/2006JD008033.
- Rodgers, C., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Singapore: World Scientific.
- Saunders, R., Matricardi, M., and Brunel, P., 1999. An improved fast radiative transfer model for assimilation of satellite radiance. *Quarterly Journal of the Royal Meteorological Society*, **125**, 1407–1425.
- Tan, D. G. H., Andersson, E., Fisher, M., and Isaksen, I., 2007. Observing-system impact assessment using a data assimilation ensemble technique: application to the ADM-Aeolus wind profiling mission. *Quarterly Journal of the Royal Meteorological Society*, **133**, 381–390.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, I., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J., 2005. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**, 2961–3012.
- Uppala, S., Dee, D., Kobayashi, S., Berrisford, P., and Simmons, A., 2008. Towards a climate data assimilation system: status update of ERA-Interim. *ECMWF Newsletter*, **115**, 12–18.
- Warner, T. T., 2011. *Numerical weather and climate prediction*. Cambridge, NY: Cambridge University Press, pp. 89–95.
- Zhu, Y., and Gelaro, R., 2008. Observation sensitivity calculations using the adjoint of the gridpoint statistical interpolation (GSI) analysis system. *Monthly Weather Review*, **136**, 335–351.

Cross-references

[Atmospheric General Circulation Models](#)
[Data Assimilation](#)
[Geophysical Retrieval, Overview](#)
[GPS, Occultation Systems](#)
[Land-Atmosphere Interactions, Evapotranspiration](#)
[Lidar Systems](#)
[Limb Sounding, Atmospheric](#)
[Microwave Radiometers](#)
[Ocean-Atmosphere Water Flux and Evaporation](#)
[Radar, Altimeters](#)
[Radar, Scatterometers](#)
[Radar, Synthetic Aperture](#)
[Radiative Transfer, Solution Techniques](#)
[Radiation, Electromagnetic](#)
[Radiative Transfer, Theory](#)
[Radio-Frequency Interference \(RFI\) in Passive Microwave Sensing](#)
[Soil Moisture](#)
[Stratospheric Ozone](#)

WETLANDS

John Melack

Department of Ecology, Evolution and Marine Biology,
University of California, Santa Barbara, CA, USA

Definitions

Passive microwave radiation is emitted from the Earth's land, seas, and atmosphere at wavelengths generally between 0.15 and 30 cm or, if expressed as frequencies, between 1 and 200 GHz.

Emissivity is the ratio of energy radiated by a material to energy radiated by a blackbody at the same temperature.

Introduction

Wetlands cover extensive areas worldwide (Lehner and Döll, 2004), have important ecological and biogeochemical functions, and play critical roles in improving water quality, mitigating floods, and providing habitat for fish and wildlife. For many wetlands, remote sensing is the preferred approach to obtain a synoptic view of inundation and vegetative cover, and a suite of optical and microwave sensing systems and analysis algorithms are being applied to wetlands (Sahagian and Melack, 1998; Melack, 2004). In the case of the large, temporally varying wetlands found throughout the world, a remote sensing system with frequent, near-global coverage and sensitivity to wetness is necessary. These requirements are met by passive microwave sensors.

Passive microwave systems and analyses

A global record of *passive microwave radiation* measured from satellites is available from 1979 to the present. The Scanning Multichannel Microwave Radiometer (SMMR) was operated on board the Nimbus-7 satellite from 1979 to

1987, with global coverage every 6 days. The Special Sensor Microwave/Imager (SSM/I) replaced SMMR in 1987 and operates today with 3 day global coverage from a satellite in the US Defense Meteorological Satellite Program. The Tropical Rainfall Measuring Mission (TRMM), launched in 1997, included a microwave radiometer similar to that on SSM/I but providing higher spatial resolution because it flies at lower altitude than the SSM/I. The Advanced Microwave Scanning Radiometer (AMSR), launched in 2002 on the Aqua satellite, offers additional capabilities.

Measurements of passive microwave radiation are expressed as brightness temperatures ($^{\circ}\text{K}$) and are recorded as vertical and horizontal polarizations at several frequencies (Choudhury, 1989). To reduce effects of atmospheric water vapor and temperature on the measurements, the difference between the two polarizations, referred to as ΔT , is often used. However, surface roughness, exposed soil and rock, seasonal vegetation changes, and other atmospheric conditions can affect the ΔT . Prigent et al. (1998) have developed an approach to calculate microwave *emissivities* of land surfaces after removing the contributions from the atmosphere, clouds, and rain and modulation by surface temperatures by using ancillary remotely sensed information and meteorological reanalyses. In general, flooded regions have low microwave emissivities and high polarization differences relative to non-flooded areas. Spatial resolutions of approximately 10–50 km limit the application of the technique to large wetlands or to regions where the cumulative area of smaller wetlands comprises a significant proportion of the landscape.

Calm water surfaces result in a strongly polarized emission at 37 GHz (e.g., SMMR ΔT ca. 60°K), although this is attenuated to varying degrees by overlying vegetation. In the absence of flooding, the dense vegetation and relatively level terrain typical of large wetlands present a stable background of depolarized microwave emission (e.g., SMMR ΔT averaging ca. 4°K). Fluctuations in the extent of inundation can be quantified if the ΔT is raised sufficiently above background. Inundation area can be estimated from the ΔT by mixing models that incorporate the microwave emission characteristics of the major landscape units (Sippel et al., 1994). The results have been validated against independent measures of flooding, such as river-stage records in areas of floodplain where inundation is known to be controlled by a large river.

Passive microwave applications

Inundation

A global monthly time series of inundation during the 1990s was produced by Prigent et al. (2007) based on a combination of passive microwave surface emissivities, scatterometer responses, and visible and near-infrared reflectances for ca. 0.25° grid cells. The detection of inundation relied primarily on SSM/I data. In forested regions it appears that the results do not indicate inundation if standing water occupies less than 10 % of the pixel.

In comparison to inundation determined under low and high water levels at 100 m resolution with synthetic aperture radar for the central Amazon (Hess et al., 2003), Prigent et al.'s results do fairly well, underestimating low water area by 11 % and high water area by 30 %. Variations in wetland area match well with variations in water level derived from TOPEX-Poseidon altimetry in the Niger, Ganges, Pantanal, and Amazon basins on a 4° grid.

In tropical areas, passive microwave remote sensing studies have focused on analysis of the inundation patterns in seasonally flooded forests and savannas and on comparative analyses of hydrological patterns in the major wetlands in South America (Hamilton et al., 2002). Observations with SMMR at the 37 GHz have been analyzed to determine spatial and temporal patterns of inundation on floodplains of the Amazon, Tocantins and Orinoco basins, and the Pantanal wetlands of South America.

Ecological and biogeochemical studies

Information about inundation and wetland vegetation are essential for the understanding of carbon dynamics in the Amazon basin. By combining field measurements of carbon dioxide concentrations in surface waters with passive and active remote sensing of inundation, Richey et al. (2002) calculated the evasion (outgassing) of CO₂ from water to the atmosphere in the central Amazon basin. Similarly, by combining measurements of methane emission from a variety of habitats and sites with inundation and vegetation variations derived from microwave and optical remote sensing analyses, Melack et al. (2004) estimated methane emissions from the Amazon basin.

Variations in the distribution and inundation of floodplain habitats play a key role in the ecology and production of many commercially important freshwater fish. In a comparison of flooded areas estimated from a monthly series of passive microwave data (Sippel et al., 1998) with the annual fish yield aggregated from the Brazilian Amazon, Melack et al. (2009) found significant relationships for small species at lower trophic levels generally at short lag times (0–1 years), while those for large species at higher trophic levels had considerably longer lag times (3–5 years).

Summary

The principal advantages of the passive microwave observations are their frequent global coverage and their ability to reveal characteristics, such as inundation, of the land surface beneath cloud cover and vegetation.

Bibliography

- Choudhury, B. J., 1989. Monitoring global land surface using Nimbus-7 37 GHz data, theory and examples. *International Journal of Remote Sensing*, **10**, 1579–1605.
- Hamilton, S. K., Sippel, S. J., and Melack, J. M., 2002. Comparison of inundation patterns among major South American floodplains. *Journal of Geophysical Research*, doi:10.1029/2000JD000306.
- Hess, L. L., Melack, J. M., Novo, E. M. L. M., Barbosa, C. C. F., and Gastil, M., 2003. Dual-season mapping of wetland inundation and vegetation for the central Amazon basin. *Remote Sensing of Environment*, **87**, 404–428.
- Lehner, B., and Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, **296**, 1–22.
- Melack, J. M., 2004. Remote sensing of tropical wetlands. In Ustin, S. (ed.), *Remote Sensing for Natural Resources Management and Environmental Monitoring*, 3rd edn. New York: Wiley. Manual of Remote Sensing, Vol. 4, pp. 319–343.
- Melack, J. M., Hess, L. L., Gastil, M., Forsberg, B. R., Hamilton, S. K., Lima, I. B. T., and Novo, E. M. L. M., 2004. Regionalization of methane emissions in the Amazon basin with microwave remote sensing. *Global Change Biology*, **10**, 530–544.
- Melack, J. M., Novo, E. M. L. M., Forsberg, B. R., Piedade, M. T. F., and Maurice, L., 2009. Floodplain ecosystem processes. In Gash, J., Keller, M., and Silva-Dias, P. (eds.), *Amazonia and Global Change*. Washington, DC: American Geophysical Union. Geophysical Monograph Series, Vol. 186, pp. 525–541, doi:10.1029/GM186.
- Prigent, C., Rossow, W. B., and Matthews, E., 1998. Global maps of microwave land surface emissivities: potential for land surface characterization. *Radio Science*, **33**, 745–751.
- Prigent, C., Papa, F., Aires, F., Rossow, W. B., and Matthews, E., 2007. Global inundation dynamics inferred from multiple satellite observations, 1993–2000. *Journal of Geophysical Research*, doi:10.1029/2006JD007847.
- Richey, J. E., Melack, J. M., Aufdenkampe, A. K., Ballester, V. M., and Hess, L., 2002. Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric carbon dioxide. *Nature*, **416**, 617–620.
- Sahagian, D., and Melack, J. M., (eds.). 1998. *Global wetland distribution and functional characterization: Trace gases and the hydrologic cycle*. International Geosphere-Biosphere Program Report 46, Stockholm, 92 p.
- Sippel, S. K., Hamilton, S. K., Melack, J. M., and Choudhury, B., 1994. Determination of inundation area in the Amazon River floodplain using the SMMR 37 GHz polarization difference. *Remote Sensing of Environment*, **48**, 70–76.
- Sippel, S. J., Hamilton, S. K., Melack, J. M., and Novo, E. M. M., 1998. Passive microwave observations of inundation area and the area/stage relation in the Amazon River floodplain. *International Journal of Remote Sensing*, **19**, 3055–3074.

Author Index

A

Abrams, Michael, 890
Albrecht, Rachel I., 339
Alpers, Werner, 433, 719
Anthes, Richard, 489
Ao, Chi O., 264

B

Bailey, G. Bryan, 320
Balk, Deborah, 869
Balstad, Roberta, 515
Barry, Roger, 104
Bauer, Peter, 912
Benediktsson, Jón Atli, 140, 503
Bennartz, Ralf, 780
Boersma, K. Folkert, 838
Boland, Stacey, 405
Brigham, Lawson, 512
Bruegge, Carol, 47

C

Cecil, Daniel J., 339
Chapron, Bertrand, 461
Colaizzi, Paul D., 291
Collard, Fabrice, 461
Colliander, Andreas, 39
Comiso, Josefino C., 727
Cooper, Steve, 70
Costa, Enrico, 219
Culshaw, Martin, 800

D

De Sherbinin, Alexander, 196
Deichmann, Uwe, 869
Desilets, Darin, 83
Diner, David J., 658, 663
Ding, Kung-Hau, 595
Dolman, Johannes A., 261
Donnellan, Andrea, 146
Doswell, Charles A., III, 767
Drinkwater, Mark, 91
Du, Yang, 150, 362, 588
Duerr, Ruth, 127
Durand, Michael, 816
Dushaw, Brian, 4

E

Eldering, Annmarie, 32
Elvidge, Christopher D., 869

England, Anthony, 558, 574
Evans, Robert G., 291
Evet, Steven R., 291

F

Farr, Thomas, 311
Fellous, Jean-Louis, 254
Ferraro, Ralph, 640
Fetzer, Eric, 909
Fisher, Joshua B., 325
Foster, James L., 821
Freeman, Anthony, 51
Friedl, Randall, 405
Fritz, Steffen, 257
Fu, Lee-Lueng, 455

G

Gabrynowicz, Joanne Irene, 332
Gaier, Todd, 186
Gail, William, 78, 159, 520
Gillespie, Alan, 303, 314
Glackin, David L., 407, 412, 691
Goodman, Steven J., 339
Guzzi, Rodolfo, 606

H

Hall, Dorothy K., 821
Hallikainen, Martti, 364
Harris, Ray, 134
Hatfield, Jerry, 22
Hemmati, Hamid, 163
Hibbitts, Charles, 846
Hook, Simon, 830
Hu, Aixue, 98
Huete, Alfredo, 788, 883
Hunsaker, Douglas J., 291
Hunt, E. Raymond Jr., 653
Hyon, Jason, 162

I

Ivins, Erik, 791

J

Johannessen, Johnny, 461

K

Kahn, Ralph, 16
Kerr, Yann, 783
Kicza, Mary, 263

Kimball, John, 886
Klatt, Calvin, 228
Komjathy, Attila, 286
Krueger, Arlin, 853, 860
Kruse, Fred A., 702
Kunkee, David, 395, 634

L

Lagerloef, Gary, 747
Lavender, Samantha, 437
Le Vine, David M., 581, 591
Lebsock, Matthew, 70
Levelt, Pieter F., 838
Liberti, Gian Luigi, 492, 498
Lin, Bing, 145
Liu, Mingyan, 190
Liu, W. Timothy, 480, 759
Livesey, Nathaniel, 344, 834
Long, David, 54, 532
Lowman, Paul, 681
Lyzenga, David R., 403

M

Macauley, Molly, 86
Marsh, Stuart, 800
Martin-Neira, Manuel, 390
Marzano, Frank S., 201, 576, 585, 624
Maslanik, James, 509
Matheou, Georgios, 35
McLaughlin, Dennis, 131
Melack, John, 921
Menzies, Robert, 334
Minnett, Peter, 101, 754
Moghaddam, Mahta, 190
Moran, Susan, 22, 88
Moussessian, Alina, 185
Muleri, Fabio, 219
Murphy, Donald L., 281
Myers, Jeffrey, 409

N

Neumann, Gregory, 821, 869
Nghiem, Son V., 821, 869

O

Oki, Taikan, 895, 903
Olsen, Nils, 358
O'Shaughnessy, Susan A., 291

P

Parsons, Mark A., 121
Perovich, Donald, 722
Perry, Sandra L., 702
Persson, Anders, 35
Pieri, David, 237
Pilkington, Mark, 232
Pinker, Rachel T., 806
Prashad, Lela, 878
Prior-Jones, Michael R., 429

Q

Qi, Jiaguo, 20

R

Rahmat-Samii, Yahya, 375, 668
Raney, Keith, 525, 536, 547
Roberts, Dar, 210
Robinson, Ian, 469
Romeiser, Roland, 426
Ruf, Christopher, 46, 389, 815

S

Salomonson, Vincent V., 684
Santee, Michelle, 796

Sapiano, Mathew R.P., 58
Shaffer, Lisa Robock, 284
Silva, Agnello, 190
Singhroy, Vernon H., 232, 328, 681
Skou, Niels, 382, 386
Small, Christopher, 869
Sorichetta, Alessandro, 869
Stammer, Detlef, 446
Stamnes, Knut, 495

T

Taylor, Mark, 35
Teixeira, Joao, 35
Thomas, Robert, 269
Thorp, Kelly, 515
Tian, Baijun, 349
Tratt, David M., 177
Tsang, Leung, 595

U

Ustinov, Eugene, 241, 247, 251

V

Van Zyl, Jakob, 136

Veefkind, J.P., 838
Velden, Chris, 849

W

Waliser, Duane, 349
Waske, Björn, 140, 503
Weaver, Ron, 119, 517
Weill, Alain, 1, 11, 13
Weng, Fuzhong, 68
Wensink, Han, 719
Willis, Josh, 743
Wilson, Cara, 202
Wood, Eric F., 58

X

Xie, Xiaosu, 480, 759

Y

Yang, Xiaojun, 73
Yarovoy, Alexander, 398
Yeh, Pat J.-F., 895, 903

Z

Zreda, Marek, 83

Subject Index

- A**
- Absolute calibration, 47
 - Absorption, 492–495
 - atmospheric scatterers, 492
 - molecular absorption, 493–495
 - propagating electromagnetic wave, 492
 - Absorption bands, 692
 - Acoustic Doppler Current Profiler (ADCP), 1, 434
 - Acoustic radiation
 - dipoles, 2–4
 - echo sounders, 1
 - energy propagation and rays, 2
 - frequency audio spectrum, 1
 - monopoles, 2–4
 - pulsing sphere, 3
 - sound amplitude, 1
 - sound propagating, 1
 - wave equation, 2
 - Acoustic sounders, 719
 - Acoustic thermometry, 8–9
 - Acoustic tomography, ocean, 4–10
 - Acoustic Thermometry of Ocean Climate (ATOC), 8
 - Acoustic turbulent scattering, 15–16
 - Acoustic waves, 11–16
 - Doppler shift propagation, 12
 - reflection and refraction, 12–13
 - Lambert scattering, 14–15
 - rigid sphere, 14
 - turbulent scattering, 15–16
 - wave absorption, 12
 - Acousto-optical tunable filter (AOTF), 695
 - Active electro-optical instruments, 697–698
 - Active Microwave Instrument (AMI), 419, 699–701
 - Active remote sensing techniques, 574
 - Active sensors, 197
 - Advanced Baseline Imager (ABI), 809
 - Advanced Microwave Scanning Radiometer (AMSR), 63
 - Advanced Microwave Scanning Radiometer on NASA's Earth Observing System (AMSR-E), 282–283, 634, 736, 737, 823
 - Advanced Microwave Sounding Unit (AMSU), 69–70
 - Advanced Scatterometer (ASCAT), 534, 535
 - Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)
 - Kliuchevskoi volcano, 892
 - land surface topography, 324
 - Mediterranean Sea, 436
 - Advanced Very High Resolution Radiometer (AVHRR)
 - aerosols, 17
 - land surface phenology (LSP), 886–887
 - Madden-Julian Oscillation (MJO), 356
 - surface radiative fluxes, 808
 - vegetation indices, 883
 - Advisory Committee on Commercial Remote Sensing (ACCRES), 81
 - Aerial photogrammetric techniques, 321
 - Aeronomy of Ice in the Mesosphere (AIM), 418
 - Aerosol Polarimetry Sensor (APS)
 - instrument, 666
 - Aerosol RObotic NETwork (AERONET), 18
 - Aerosols, 16–20, 660
 - Advanced Very High Resolution Radiometer (AVHRR), 17
 - Aerosol RObotic NETwork (AERONET), 18
 - Along Track Scanning Radiometer (ATSR), 17
 - climate and health impact, 16
 - elastic scattering, 498–501
 - environmental effects, 19
 - French Space Agency (CNES), 18
 - health and direct sampling applications, 18
 - human activities, 16
 - Moderate resolution Imaging Spectroradiometer (MODIS), 18
 - Multi-angle Imaging Spectroradiometer (MISR), 18
 - physical and chemical properties, 17
 - POLarization and Directionality of Earth's Reflectance (POLDER), 18
 - satellite observations, 18
 - solid/liquid particles, 16
 - Stratospheric Aerosol and Gas Experiment (SAGE) instruments, 17
 - surface-based sun photometry, 17
 - surface lidar networks, 18
 - Total Ozone Mapping Spectrometers (TOMS), 17
 - Agricultural expansion and abandonment, 20–22
 - Agriculture and remote sensing, 22–30
 - crop biomass, 27
 - crop ground cover, 27–28
 - crop nutrient status, 29
 - crop surfaces, 23
 - crop yield, 28
 - intercepted solar radiation, 27
 - leaf chlorophyll content, 26–27
 - leaf reflectance, 23
 - pests and potential stresses, 29–30
 - spectrophotometers, 23
 - Stefan-Boltzmann relationship, 23
 - thermal radiation, 28–29
 - vegetative indices (VIs), 23–26
 - Agulhas Current, 466–468
 - Airborne Multiangle SpectroPolarimetric Imager (AirMSPI), 665
 - Airborne Multi-Spectral Sunphoto-and Polarimeter (AMSSP), 664–665
 - Airborne observational platforms, 409–411
 - airborne radar platforms, 411
 - air sampling platforms, 411
 - cloud cover, 409
 - data systems, 411
 - manned aircraft, 409–410
 - mechanical vibration, 411
 - remote sensing aircraft, 410–411
 - thermal infrared channels, 411
 - unmanned aerial vehicles (UAVs), 410
 - Airborne Optical Spotlight System-Multispectral Polarimeter (AROSS-MSP), 665
 - Airborne Polarization and Directionality of Earth's Reflectances (AirPOLDER)
 - instrument, 664
 - Airborne radar platforms, 411
 - Airborne SAR system (AIRSAR), 52
 - Airborne sensors, 831–832
 - Air pollution, 32–34
 - Earth's atmosphere, 32
 - gases/particles concentration, 32
 - Interferometric Monitoring of Greenhouse Gases (IMG), 32
 - Measurement of Air Pollution from Satellites (MAPS) experiment, 32
 - Measurements of Pollution in the Troposphere (MOPITT) instrument, 32
 - NO₂, 33
 - SO₂, 33–34
 - Total Ozone Mapping Spectrometer (TOMS), 33
 - tropospheric CO, 33
 - tropospheric ozone, 33
 - Air sampling platforms, 411
 - Alaska Volcano Observatory (AVO), 890, 891
 - Albedo, 225, 226, 788, 806
 - Along-track interferometry (ATI), 426–428
 - Along-Track Scanning Radiometer (ATSR)–1, 659

- Alpha-derived emissivity (ADE) method, 309
 Alpha-residual method, 307–308
 Altimeters, 525–530, 700
 flight systems, 528
 ice-sheet radar altimetry, 528–529
 measurement precision, 527, 528
 mesoscale features, 526
 near-polar ice, 527
 oceanographic vs. geodetic orbits, 526–527
 orbit considerations, 529–530
 propagation delay corrections, 527
 pulse-limited geometry, 527, 528, 530
 satellite-based radar altimeter, 530–531
 sea-surface corrections, 527
 sea-surface height (SSH), 526
 significant wave height (SWH), 526
 Skylab, 525
 wind speed (WS), 526
 Amplified spontaneous emission (ASE), 168
 Amundsen Sea, 272
 Angular dependent model (ADM), 146
 Angular reflectances, 661
 Antarctica ice sheet
 Arctic-wide glacier recession, 269
 Circumpolar Current, 270
 Drake's Passage, 270, 271
 East and West Antarctica, 270
 Geoscience Laser Altimeter System (GLAS), 275
 global sea level, 269
 Gravity Recovery and Climate Experiment (GRACE), 275–276
 grounding line, 270
 ICESat, 275
 ice shelves, 270, 272
 Last Glacial Maximum, 270
 mass balance, 277–280
 passive-microwave radiometer, 274, 275
 radar altimeter, 274–275
 scatterometer, 274
 snowfall, 278
 surface elevation, 275, 276
 synthetic aperture radar (SAR), 274
 Transantarctic Mountains, 270
 West Antarctica ice sheet (WAIS), 272
 Antarctic ozone hole, 834
 Antarctic sea ice, 728
 Antenna effects, 46
 Antenna temperature, 383
 Anvil cloud, 768, 769
 Aperture blockage, 679
 Aperture field (AF) method, 668
 Aperture synthesis radiometry, 391–394, 699
 calibration, 394
 collection of N antennas, 392
 Corbella equation, 391, 393
 field of view, 393
 flat target transformation, 393–394
 historical development, 391
 polarimetry, 394
 radiometric sensitivity, 393
 spatial frequency sampling and aliases zones, 393
 spatial resolution, 391–393
 spatial windowing, 392–393
 Van Cittert–Zernike theorem, 392
 Appleton-Hartree formula, 286, 287
 Appleton-Hartree magnetoionic theory, 286–287
 Aquaculture, 207
 Aquarius/SAC-D mission, 752
 Arago, 663
 Arakawa C-grid, 448, 449
 Arctic region, 727
 Arctic-wide glacier recession, 269
 1999 Arctic Climate Observations Using Underwater Sound (ACOUS) experiments, 8
 Argo Project, 744
 ARGOS, 430, 432
 Asia-Pacific Satellite Data Exchange and Utilization Group (APSDEU), 263
 Atmospheric correction, 17
 Atmospheric Dynamics Mission (ADM)
 top-of-the-atmosphere radiation, 146
 tropospheric winds, 418
 Atmospheric general circulation models, 35–37
 climate prediction models, 36
 fluid dynamics and thermodynamics equations, 35
 model dynamics, 36
 model physics, 36–37
 weather prediction, 36
 Atmospheric motion vectors (AMVs), 849–850
 Atmospheric opacity, 558
 Atmospheric scattering, 854
 Atmospheric UV absorbers, 853–854
 Atmospheric water vapor, 755
 Austin Post aerial photograph, 104
 Autocorrelation spectrometers, 389–390
 Averaging kernel, 250
 AVHRR NDVI, 656
 Ayles Ice Shelf, on Ellesmere Island, Canada, 112, 281
 Azimuth processing, 537–538
 Azimuth resolution, 538, 544, 555
B
 Babinet, 663
 Backscatter calibration, 55–56
 Backscatter lidar, 697
 Backscatter ultraviolet (BUV) instrument, 861, 862
 Bare earth model, 323
 Baseline decorrelation, 540
 Basic Plan for Space Policy, 81
 Basic Space Law in 2008, 81
 Bathymetry Assessment System (BAS), 721
 Bayh-Dole Act of 1980, 490–491
 Beaufort gyre, 732
 Beijing case study, 869–876
 Bidirectional reflectance, 403–405
 Bidirectional reflectance distribution function (BRDF), 317, 619, 658, 789
 albedo, 788
 Lambert surface, 497
 reflected radiance, 497
 sea ice albedo, 722–723
 soil optical properties, 788
 Biomass, 214
 Bi-spectral methods, 641
 Blue Planet, 896
 Bookend vortices, 775
 Bootstrap algorithm, 109, 730–731
 Born approximation, 15
 Boundary conditions, 151
 Bow echo morphology, 771
 Bragg scattering, 15, 551
 Bragg's law, 224
 Brewer-Dobson circulation, 797
 Brewster, 663
 Brightness temperature, 730
 British Antarctic Survey, 112
 Bruggeman–Hanai–Sen mixing model, 399
C
Calanus finmarchicus, 207
 Calibration
 lidars, 698
 microwave radiometers, 46–47, 699
 optical/infrared passive sensors, 47–51
 scatterometers, 54–56
 synthetic aperture radars (SAR), 51–53
 Calibration and validation, 39–43
 climatic temporal context, 42
 Committee on Earth Observation Satellites (CEOS), 40
 geophysical products, 41
 international cooperation, 40
 NASA's Nimbus-7 satellite in 1978, 40
 prelaunch phase, 42
 remote sensing products, 39–40
 satellite remote sensing era, 40
 sensor products, 40–41
 spatial scaling, 41–42
 validation stage, 42–43
 CALIOP lidar, 72
 Canadian Space Agency Act of 1990, 81
 Canopy fuels, 213, 215
 Cassegrain reflector system, 677, 679
 Cassiopeia A, 582
 Catch per unit effort (CPUE), 202
 CHALLENGING Minisatellite Payload (CHAMP), 419
 Chimney clouds, 646
 Chlorophyll-a (Chl-a), 438–441
 Circular horn, 375, 376
 Circularly polarized wave, 579
 Circumpolar Current, 270
 Climate data records (CDR), 58
 Climate monitoring and prediction, 58–66
 CPC temperature probability product, 66
 ozone, 60–61
 precipitation, 61
 reanalysis data, 63–65
 remotely sensed observations, 59
 sea ice, 62–63
 seasonal climate forecasts, 65
 sea surface temperature, 63, 64
 statistical and dynamical prediction models, 65–66
 top of the atmosphere (TOA) radiation budget, 59–60
 vegetation, 61–62
 weather scales, 59
 Climate Prediction Center Merged Analysis of Precipitation (CMAP), 61, 647–649
 Clay soils, 802
 Cloud Absorption Radiometer (CAR), 659
 Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO), 18, 336, 415, 416, 418, 689, 809
 Cloud condensation nuclei (CCN), 19
 Cloud Imaging and Particle Size Experiment (CIPS), 660
 Cloud indexing methods, 640–641
 Cloud liquid water, 68–70
 Advanced Microwave Sounding Unit (AMSU), 70
 atmospheric transmittance, 69
 brightness temperatures, 69
 hydrometeor variables, 68
 intertropical convergence zone (ITCZ), 69, 70
 Nimbus-7 scanning multichannel microwave radiometer (SMMR) data, 68
 optical thicknesses, 69
 passive and active remote sensing technology, 69
 Special Sensor Microwave Imager (SSM/I), 68
 thermal emission, 68
 Cloud model techniques, 641
 Cloud Profiling Radar (CPR), 646
 Cloud properties, 70–73
 CALIOP lidar, 72
 CloudSat cloud profiling radar, 72
 droplet size distribution, 71

- infrared techniques, 71–72
 microwave techniques, 71
 reflection and absorption properties, 70–71
 turbulent and microphysical processes, 71
 visible and near-infrared techniques, 71
 visible polarization techniques, 72
 Cloud-radiative forcing, 807
 Clouds, 660
 Clouds and the Earth's Radiant Energy System (CERES), 146
 CloudSat, 418, 646, 647
 Coastal ecosystems, 73–78
 aquatic flowering plants, 74
 benthic habitats mapping, 77
 benthic system, 74
 coastal marshes, 74
 conventional field-based mapping methods, 74
 estuary, 74
 mangrove ecosystem, 74
 ocean color radiometers, 77
 quality indicators, 77
 spatial complexity and temporal variability, 74
 submerged aquatic vegetation (SAV) mapping, 77, 78
 terrestrial and oceanic ecosystems, 73
 terrestrial/open-ocean applications, 75–77
 terrestrial sensors, 77
 water quality measures, 77–78
 watershed landscape characterization, 77
 wetland mapping, 77
 Coastal Zone Color Scanner (CZCS), 208, 437
 Coherence, 590–591
 Coherency time, 580
 Coherent superposition, 603, 604
 Cole–Cole model, 400
 Commercial remote sensing, 78–81
 aerial remote sensing, 79
 AstroVision, 80
 commercial market, 78–79
 COSMO-SkyMed, 82
 Disaster Monitoring Constellation (DMC), 80
 EarthWatch, 79
 FORMOSAT, 82
 GeoEye–1, 81
 GIS and photogrammetry, 81
 Indian Remote Sensing (IRS), 82
 Korean KOMPSAT–1/2 satellites, 82
 Land Remote Sensing Commercialization Act of 1984, 79
 Landsat Data Continuity Mission, 79–80
 legislation, 80
 licenses, 81
 mosaics, 79
 NextView contracts, 79
 optical photogrammetry, 79
 radar and LIDAR, 81
 radar imagery, 82
 RapidEye, 80
 regulatory framework, 81
 Seaviewing Wide Field-of-view Sensor (SeaWiFS) mission, 80
 spaceborne commercial remote sensing, 79
 spaceborne remote sensing technologies, 80
 Space Imaging, 79
 Spot Image, 79, 81–82
 Surrey Satellite Technology Ltd. (SSTL), 80, 82
 TerraSAR satellite system, 80
 Commercial Remote Sensing Policy of 2003, 81
 Committee on Data Management and Computation (CODMAC), 518
 Committee on Earth Observation Satellites (CEOS), 40, 121, 263, 520
 Compact High Resolution Imaging Spectrometer (CHRIS), 660
 Compact/hybrid polarimetry, 53
 Compton scattering, 221
 Condensation nuclei, 768
 Conductivity-temperature-depth (CTD), 434
 Conical horns, 379, 380
 Consensus theory, 142–143
 Continuum radiation, 582–584
 Contour (shaped) beams, 668
 Conventional microwave radiometer, 386–389
 conical scan, 388
 Dicke radiometer, 386, 387
 hybrid radiometer, 387–388
 implementation, 388
 noise injection radiometer (NIR), 386–387
 push-broom radiometer system, 388–389
 single channel receiver, 386, 387
 total power radiometer, 386, 387
 Coordination group for meteorological satellites (CGMS), 263
 Convective storms, 769–773, 778
 Copepods, 207
 Coral reef system, 205
 Corbella equation, 391, 393
 Coriolis satellite, 420
 Correlation microwave radiometers, 389–390
 Corrugated horns, 380–381
 Cosmic microwave background, 581
 Cosmic-ray hydrometeorology, soil water, 83–86
 atmospheric collision, 86
 cosmic-ray neutrons, 83–84
 neutron intensity, 85
 neutron interactions in soil, 83
 soil moisture, 84–85
 Cost benefit assessment, 86–88
 Crop biomass, 27
 Crop coefficient, 326
 Crop stress, 88–90
 crop physiological adaptations, 89
 digital elevation models (DEM), 90
 environmental factors, 88
 growth and development, 89
 herbicide applications, 89
 leaf expansion and senescence, 89
 leaf stomata, 88
 lush cotton growth, 90
 pests, 89
 profitable yield and minimizing environmental impact, 88
 weed interference, 89
 Crop water stress index (CWSI), 29, 293–296
 Crop yield, 28
 Cross-track interferometry (XTI), 426
 CryoSat altimeter, 527, 529
 Cryosphere
 aerial photography, glaciers and ice caps, 104–105
 climate change effects, 98–101
 climate change feedbacks, 101–104
 freshwater ice, 108
 frozen ground, 113
 glacier mapping, 105–106
 glacio-isostatic adjustment, 148–149
 icebergs, 113
 ice sheet dynamics, 149
 ice sheet remote sensing, 111–112
 ice shelves, 112
 satellite imagery and data, 105
 sea ice, 108–111, 149
 snow cover, 106–108
 Cryosphere observing system (CryOS), 91–97
 albedo and surface temperature measurements, 96
 climate variability and change, 92
 freshwater resources, 91
 gravity and geoid, 96
 ground-based observations, 96
 marine cryosphere, 92
 measurement reference frames, 97
 permafrost, 93–94
 remote sensing product validation, 97
 river and lake ice, 93
 satellite communications and positioning, 96–97
 satellite elements, 92
 sea level rise prediction, 91
 snow extent and water equivalent, 92–93
 solid precipitation, 93
 terrestrial cryosphere, 92
 weather and climate prediction, 91
 Current best estimate (CBE), 752
 Cyclonic vortex, 775
 Cygnus A, 582, 584
D
 Data access, 119–121
 Committee on Earth Observation Satellites (CEOS), 121
 Compact Disk Read-Only Memory (CD-ROM), 119
 data delivery protocols, 121
 data discovery, 120
 data information technology systems, 120
 data management system, 120
 digital data, 119
 file transfer protocol (FTP), 119, 120
 Global Change Master Directory, 121
 ingest processes, 120
 online data storage, 120
 Open Data Access Protocol (OpenDAP), 121
 secure file transfer protocol (SFTP), 120
 Thematic Realtime Environmental Distributed Data Services (THREDDS) system, 121
 Data archival and distribution, 122–126
 digital data, 122
 floppy disk, 122
 life cycle, 122
 OAIS Reference Model, 123–126
 Data archives and repositories landscape, 127–130
 international systems, 128–129
 national centers and systems, 129
 research data, 130
 state and local agencies, 129–130
 Data assimilation, 131–133
 Data Collection Service (DCS), 430
 Data policies, 134–136
 European Union, 135
 Global Earth Observation System of Systems (GEOSS), 136
 security, 135–136
 United Nations (UN), 134–135
 United States, 135
 World Meteorological Organization (WMO), 135
 Data processing, SAR sensors, 136–140
 active remote sensing instruments, 136
 azimuth compression, 140
 Doppler bandwidth, 139
 flight path, 137
 imaging radars, 137
 integration time, 139
 pulse repetition frequency (PRF), 139
 radar resolution, 137–138
 range compression, 139
 range curvature correction, 140
 range walk, 140

- rectangular imaging algorithm, 139
 - reflected signals, 137, 138
 - Taylor series expansion, 139
 - DAWN mission, 227
 - Death Valley, California, 304, 305
 - Debye-Cole dual-dispersion mode, 373
 - Debye equation, 747
 - Debye-like model, 369
 - Debye model, 366, 367
 - Debye process, 566–568
 - Decision fusion, classification of multisource data, 140–143
 - data-level fusion, 141
 - decision-level fusion, 141
 - Earth Observation (EO) systems, 140
 - feature-level fusion, 141
 - greater quality, 141
 - land cover classification, 141–143
 - multispectral and SAR imagery, 140
 - Decision trees (DT), 505–506
 - Deconvolution method (DCM), 872
 - Deep convective storms, 768–770
 - Deep ocean convection, 727
 - Defense Meteorological Satellite Program (DMSP), 282, 686, 871
 - Degree of circular polarization (DOCP), 663, 665
 - Degree of linear polarization (DOLP), 663–665
 - Denitrification, 799
 - Dense random distribution, 585, 586
 - Dense sampling method (DSM), 869, 872–875
 - Derechos, 771
 - Dicke radiometer, 384
 - Dielectric horns, 382
 - Difference Normalized Burn Ratio (DNBR), 216
 - Differential absorption lidar (DIAL), 180, 697
 - Differential scattering coefficient, 627, 628
 - Diffraction Optical Element (DOE), 335
 - Diffusive scattering, 586
 - Digital canopy model (DCM), 213
 - Digital Photometric Mapper (DPM), 664
 - Digital tectonic activity map of earth, 681, 682
 - Dipolar momentum, 3
 - Direct ray, 593
 - Disaster Monitoring Constellation (DMC) satellite, 80, 82, 422
 - Discrete-ordinate eigen-analysis technique, 632
 - Discrete ordinate method, 618–619
 - Division-of-amplitude polarimeters, 663
 - Division-of-aperture polarimeters, 663
 - Doppler effect, 580
 - Doppler frequency shift, 580
 - Doppler lidar, 180–181, 697–698
 - Doppler shift, 12
 - Drake's Passage, 270, 271
 - Dry adiabatic lapse rate, 768
 - Dwell time per footprint, 388
- E**
- Earth Clouds, Aerosols and Radiation Explorer (EarthCARE), 418, 646
 - Earth crust, 147–148
 - earthquakes, 147–148
 - plate tectonics, 147
 - volcanoes, 148
 - Earth Exploration Satellite Service (EESS), 639
 - Earth Observation Satellite Corporation (EOSAT), 522
 - Earth Observing System's (EOS), 417, 418, 688–690
 - Earth Radiation Budget Experiment (ERBE), 146
 - Earth Radiation Budget Satellite (ERBS), 418
 - Earth radiation budget, top-of-atmosphere radiation, 145–146
 - Earth Resources Technology Satellite (ERTS), 105
 - Earth's atmosphere
 - albedo, 225, 226
 - attenuation coefficient, 220
 - Compton scattering, 221
 - cosmic rays, 225
 - shower, 225–226
 - terrestrial gamma-ray flashes (TGFs), 226
 - Earth's core, 147
 - Earth's mantle, 147
 - Earth–Sun distance, 806
 - Earth system models, 146–149
 - core, 147
 - cryosphere, 148–149
 - lithosphere and crust, 147–148
 - mantle, 147
 - Echo sounders, 1
 - Ecosystem approach to fisheries (EAF), 204
 - Ecosystem degradation, 655
 - Eddy available potential energy (EAPE), 353
 - Edge taper (ET), 669–670
 - Effective overburden pressure, 802
 - Elastic backscatter lidar, 179–180
 - Elastic scattering, 498–501
 - aerosols, 498–501
 - hydrometeors, 498–501
 - Electric polarization, 566–568
 - Electromagnetic radiation, 576–581
 - amplitude, 579
 - coherency, 580
 - Doppler effect, 580
 - electromagnetic waves, 577
 - free-space radiation, 578
 - generation and detection, 580–581
 - Maxwell's differential equations, 577–578
 - phase, 579
 - polarization, 579–580
 - quantities, 579
 - quantum radiation, 578–579
 - spectrum, 577
 - Electromagnetic wave propagation, 150–157
 - boundary conditions, 151
 - Brewster angle, 153
 - cylinders, 156–157
 - enhanced advanced IEM (EAIEM), 153
 - Green's function, 152–153
 - integral equation method (IEM), 152
 - Kirchhoff approximation (KA), 152
 - Maxwell equations, 150–151
 - plane waves, 152
 - Poynting's theorem, 152
 - probability density function, 153
 - rough surface scattering models, 152
 - scalar and vector potentials, 151
 - slope statistical Kirchhoff (SSK) model, 153
 - small perturbation method (SPM), 152
 - statistical IEM (SIEM) model, 154
 - unifying methods, 152
 - vegetation canopy, 155–156
 - Electro-optical (E-O) spectrum, 692–693
 - Ellesmere Island, Arctic Canada, 112
 - Elliptically polarized wave, 579
 - Emerging applications, 159–162
 - anticipated applications, 161–162
 - business efficiency, 160
 - business sector, 159
 - consumer market, 159
 - consumer products and service, 160
 - ecosystem services, 160
 - public sector, 159
 - Emerging technology, 162–163
 - free-space optical communications, 163–175
 - lidar, 177–182
 - radar, 185–186
 - radiometer, 186–190
 - sensor web, 190–195
 - Emission, 495
 - Emission-based methods, 643
 - Emissivity, 211
 - Enhanced advanced integral equation method (EAIEM), 154
 - Environmental Satellite (Envisat), 417–418
 - Environmental Science Services Administration (ESSA), 350
 - Environmental treaties, 196–199
 - commercial software packages, 197
 - computer processing power, 197
 - constraints, 197–198
 - Global Earth Observation System of Systems (GEOSS), 197
 - Intergovernmental Panel on Climate Change (IPCC), 196
 - international environmental law community, 196
 - Millennium Ecosystem Assessment, 196
 - political support, 197
 - Envisat ASAR Wide Swath image, 435, 436
 - Equal-Area Scalable Earth (EASE) Grid format, 111
 - Equatorial cold tongue, 349
 - ERS-1 and-2 Advanced Microwave Instrument scatterometer mode (ESCAT), 534, 535
 - ERS-1 SAR image
 - Waddensee, 721
 - Xinchuan Gang Shoal, 721
 - ERS-2 SAR image, 435
 - Estuaries, 74
 - European Aerosol Research Lidar Network (EARLINet), 18
 - European Centre for Medium Range Weather Forecasting (ECMWF) forecast, 916
 - European Space Agency (ESA), 61
 - Evapotranspiration (ET), 325–328
 - ecophysiological constraint functions, 327
 - eddy covariance technique, 325
 - empirical, semiempirical approaches, 325–326
 - energy balance, 326
 - land surface models/earth system models, 327
 - open water evaporation and plant transpiration, 325
 - Penman-Monteith equation, 326
 - PT-JPL product, 327
 - water balance equation, 326
 - Extended boundary condition method (EBCM), 156–157
 - Extreme ultraviolet (EUV), 857–859, 867
- F**
- Faraday's law, 151
 - Far infrared (FIR) spectrum, 692
 - Far ultraviolet (FUV)
 - remote sensing, 857–859
 - sensors, 867
 - Federation of Astronomical and Geophysical data analysis Services (FAGS), 128
 - Federation Satellite-1 (FEDSAT-1), 419
 - Feed taper (FT), 669–670
 - Femtosecond white-light lidar, 178
 - Field-programmable gate arrays (FPGAs), 190
 - Fields and radiation, 201
 - Field surface energy balance (FSEB), 293, 295, 297
 - Filchner-Ronne Ice Shelf, 272
 - Fire radiative power (FRP), 215
 - First-order multiple scattering, 586
 - Fisheries, 202–208
 - anthropogenic effects, 202
 - aquaculture species, 207
 - catch per unit effort (CPUE), 202

- and climate, 207–208
coral reef monitoring, 205
electronic tagging, LMRs, 205
food supply, 202
food web models, 203, 204
habitat assessments, 202
International Council for the Exploration of the Seas (ICES), 202
living marine resources (LMRs), 202
monk seals, 206
ocean food web, 203
operational fisheries (harvesting, 203–204
right whales, 206–207
satellite ocean color, 203
sea surface temperature (SST), 203
sea turtles, 207
stock assessment, 202, 204–205
survey support, 205
transition zone chlorophyll front (TZCF), 208
vessel monitoring system (VMS), 206
Flat target response (FTR), 393–394
Fluorescence lidar, 698
Folded optics system, 676–679
 Cassegrain reflector system, 677, 679
 front-fed parabolic-reflector antenna, 676, 679
 Gregorian reflector system, 678, 679
Foldy-Lax multiple scattering equations, 598, 603
Forestry, 211–217
 active remote sensing systems, 211–212
 deforestation, 216, 217
 discriminate tree species, 211, 212
 emissivity, 211
 forest degradation, 217
 forest health, 216–217
 forest inventories, 212–214
 land-cover mapping, 217
 leaf area index (LAI), 211
 passive remote sensing systems, 212
 wildfire fuels and forest fires, 214–216
Fossil water, 903
Fourier series expansions, 609–611
Fourier transform, 390
Fram Strait, 7
Frank rock avalanche, 331
Frank Slide, Canada, 331
Frazil ice, 728
Free-space optical communications, 163–175
 communications link design, 164–165
 direct detection and coherent heterodyne reception, 165
 earth-based aperture, 173, 174
 end-to-end link block, 164, 165
 flight transceiver, 169–171
 high-capacity communications, 164
 homodyne techniques/heterodyne techniques, 166
 hybrid laser communications and precision laser ranging, 173–175
 optical channel (atmospherics), 172–173
 optical link, 164
 phase fluctuations, 166
 photon-starved regime, 165
 pointing, acquisition, and tracking (PAT), 171–172
 pulse-position modulation (PPM), 166–169
 turbulence-constrained system, 166
Free-space radiation, 578
Freezing nucleus, 773
French Alps, 106
French-Belgian coast, 720
French Space Agency (CNES), 18
Frequency audio spectrum, 1
Fresnel surface, 626
Fresnel reflection, 152, 747–748
Fried's atmospheric coherent length, 166
Front-fed parabolic-reflector antenna, 676, 679
Frozen ground, 113
Full-wave multiple scattering, 603–605
Fundamental climate data records (FCDR), 58
G
Galactic radiation, 581–583
Gamma and X-radiation, 219–227
 asteroid, 227
 Bragg's law, 224
 coded masks, 224
 crystal lattices, 219
 DAWN mission, 227
 diffraction, 224
 gamma-ray spectrometer (GRS), 226
 high energy radiation, 223
 ionizing electrons, 221–222
 Japanese satellite Kaguya/SELENE
 operative, 227
 light detectors, 222
 lobster eye optics, 224–225
 Mercury Planetary Orbiter, 227
 multilayer optics, 224
 nuclear activity, 219
 planetary probes, 227
 reflectivity for gold, 223
 units of electron volt (eV), 219
 X-ray telescopes, 223
Gamma-ray region, 577
Gamma-ray spectrometer (GRS), 226
Gas correlation spectroradiometer (GCS), 695
Gauss-Seidel method, 613–615
General circulation models (GCM), 35–37
General multipole technique (GMT), 157
GEO birds, 416, 692
Geobotany, 789
Geodesy, 228–231
 Earth's orientation, 230
 Global Geodetic Observing System (GGOS), 229
 Global Navigation Satellite System (GNSS), 229
 International Association of Geodesy (IAG), 228
 International Terrestrial Reference Frame (ITRF), 229
 orientation and gravity field, 228
 physical geodesy, 230–231
 reference frames, 229–230
Geodetic signals, 526
GeoEye–1, 81
Geological mapping using Earth's magnetic field, 232–237
 analytic signal maps, 235
 contact maps, 234
 induced magnetization, 232
 interpretation methods, 233
 interpreting magnetic field maps, 235–237
 magnetic and radar fused image, 235
 magnetic field maps, 234
 magnetic metamorphic and igneous lithologies, 232, 233
 magnetic vertical gradient maps, 234
 remanent magnetization, 232
 shaded relief magnetic maps, 235
 tilt maps, 234–235
 titanomagnetites, 232–233
Geometrical optics (GO) method, 153, 668
Geometrical theory of diffraction (GTD), 668
Geometric calibration, 697
Geomorphology, 237–240
Geophysical model function (GMF) calibration, 55–57, 533–534
Geophysical retrieval
 forward models, 241–247, 251–253
 inverse problems, 247–251, 253
 principal components, 254
Geosat orbit, 529
Geoscience Laser Altimeter System (GLAS), 275
Geostationary Earth Radiation Budget (GERB), 418, 808
Geostationary Operational Environmental Satellite (GOES), 686, 808
 lightning, 343
 surface radiative fluxes, 808
Geostationary satellite
 Data Collection Service (DCS), 430
 Inmarsat's constellation, 429–430
 Thuraya, 430
 very small aperture terminal (VSAT), 430
Geostrophic balance, 455
German Act on Satellite Data Security, 334
German TerraSAR-X satellite, 423
Gini index, 505
Glacial isostatic adjustment (GIA), 791
Global Atmosphere Watch (GAW), 128
Global Atmospheric Research Program (GARF), 518
Global Change Master Directory, 121
Global Change Observation Mission-A1 (GCOM-A1), 418
Global climate observing system (GCOS), 254–257, 263
 climate data and information, 254
 climate monitoring principles, 255
 essential climate variables, 255
 global climate observing system, 255–257
 Steering Committee, 255
 World Climate Programme, 254
Global Earth Observation System of Systems (GEOSS), 121, 136, 197, 257–260
 data standards and data dissemination, 259
 Earth observation systems, 258
 Group on Earth Observation (GEO), 258
 implementation, 258–259
 origin, 258
 societal benefit areas (SBAs), 258–261
Global Energy and Water Cycle Experiment (GEWEX), 327–328
Global Geodetic Observing System (GGOS), 229
Global Information and Early Warning System (GIEWS), 128
Global Land Ice Measurements from Space (GLIMS), 105
Global land observing system, 261–262
Global Navigation Satellite System (GNSS), 229
Global Ozone Monitoring Experiment (GOME), 61, 665
Global Ozone Monitoring Instrument (GOME), 856–857, 862–863
Global Positioning System (GPS), occultation systems, 264–267
Global Precipitation Climatology Project (GPCP), 61, 352, 647–649
Global Precipitation Measurement (GPM), 418, 646
Global programs, operational systems, 263–264
 Asia-Pacific satellite data exchange and utilization group (APSEU), 263
 Committee on Earth observation satellites (CEOS), 263
 Coordination group for meteorological satellites (CGMS), 263
 Global climate observing system (GCOS), 263
 Group on Earth observations (GEO), 264
 International Charter "Space and Major Disasters," 264

- North America-Europe data exchange group (NADEX), 264
 World meteorological organization programme (WMOSP), 264
 Global space weather systems, 857, 859
 Globalstar, 433
 Glory NASA satellite, 418
 Grabens, 148
 Gravimetry, 231
 Gravity Recovery and Climate Experiment (GRACE)
 Earth's gravitational field, 423
 Earth's gravity, 231
 ice caps and isolated glaciers, 277
 ice sheet, 63
 sea level rise, 745
 solid earth mass transport, 792–793
 surface water, 818
 Graybody emissivity method, 308
 Greenhouse gas, 909
 Greenland and Antarctic ice sheets, 100
 Greenland ice sheet, 149
 Geoscience Laser Altimeter System (GLAS), 275
 glaciers, 272, 273
 icebergs, 270
 ice cores, 273
 ice shelves, 272
 ice velocity, 270, 271
 Jakobshavn Glacier, 279
 Little Ice Age, 269
 mass balance, 277–280
 ocean warming, 279
 radar altimeter, 274–275
 satellite passive-microwave data, 274, 275
 scatterometer, 274
 sealevel rise, 272
 surface elevation, 275, 276, 279
 synthetic aperture radar (SAR), 274
 warming climate, 279
 Greenland Sea Project, 7
 Green's function, 151–152
 Green water, 903
 Gregorian reflector system, 678, 679
 Gross primary productivity (GPP), 27
 Ground-based remote sensing, 861
 Ground-penetrating radar (GPR), 401
 Group on Earth observations (GEO), 264
 Gust front, 769
- H**
 Hailstone, 772–775
 Haiti earthquake, 329, 330
 Half-space Green's function, 603
 Heard Island Feasibility Test (HIFT), 8
 Heat stress, 878, 880
 Helmholtz wave equation, 572
 Helsinki University of Technology (HUT–2D), 391
 High-electron-mobility transistor (HEMT), 188
 High-resolution precipitation products (HRPPs), 649–650
 High spectral resolution lidars (HSRL), 336
 Homogeneous nucleation temperature, 773
 Hudson Bay, 108
 Hudson Strait, 108
 Hybrid polarization, 541
 Hybrid radiometer, 387–388
 Hydrometeors, elastic scattering
 chemical composition, 499
 condensed water vapor, 498
 physical parameters, 499
 single-scattering optical properties (SSOP), 499–501
 Hyperspectral imagery (HSI), 693–695
 Hyperspectral imaging polarimeter (HIP), 665
- I**
 Ice-albedo feedback effect, 727
 Icebergs, 281–284
 active microwave systems, 283
 mass balance and freshwater flux, 281
 oil platforms, 281
 passive microwave systems, 282–283
 properties and sizes, 281–282
 remote sensing information, 282
 RMS Titanic struck, 281
 US Coast Guard International Ice Patrol (IIP), 281
 Ice, Cloud, and land Elevation Satellite (ICESat), 277
 Ice Patrol, Gulf of St. Lawrence, 113
 Ice rumples, 272
 Ice sheet, 149
 Ice sheet and ice volume, 269–281
 Antarctica, 269
 greenhouse gas concentrations, 269
 Greenland, 269
 ice cores, 273–274
 ICESat, 275
 ice streams, 270, 271
 internal deformation, 269–270
 mass balance, polar ice sheets, 276–277
 outlet glaciers-ice-shelf breakup, 277–278, 280
 Pine Island Glacier, 279
 steady state, 276
 surface mass balance, 269
 total mass balance, 269
 Ice-sheet radar altimetry, 528–529
 Ice shelves, 270, 271
 Ice streams, 270, 271
 Ice surface roughness, 661
 Import vector machines (IVMs), 142
 Improved IEM model (I-IEM), 153
 Indian National Center for Ocean Information Services (INCOIS), 204
 Indian Remote Sensing (IRS), 82
 India's IRS–2B/Oceansat–2, 420
 India's IRS-P4/Oceansat–1, 420
 Indo-Pacific warm pool, 349
 Induced magnetization, 232, 359
 Inelastic backscatter lidar, 179
 Initial boundary value problem, 607–609
 Integrodifferential equation, 629
 Interactive Multisensor Snow and Ice Mapping System (IMS), 62
 Interferogram, 322
 Interferometers microwave radiometers, 390–394
 Interferometric baseline, 53
 Interferometric synthetic aperture radar (InSAR), 53
 Earth's surface, 148
 landslides, 330–332
 land surface topography, 322
 subsidence, 804
 Intergovernmental Panel on Climate Change (IPCC), 100, 101, 808
 Internal deformation, 269–270
 Internal solitons/internal solitary waves (ISWs), 434
 Internal waves, 433–436
 acoustic Doppler current profiler (ADCP), 434
 Andaman Sea, 434–435
 conductivity-temperature-depth (CTD), 434
 internal solitons/internal solitary waves (ISWs), 434
 Mediterranean Sea, 436
 optical sensors flying detector, 433
 sea surface elevation, 433
 South China Sea, 435–436
 Strait of Gibraltar, 433
 synthetic aperture radar (SAR), 433
 thermocline, 434
 tidal flow, 434
 International Association of Geodesy (IAG), 228
 International Celestial Reference Frame (ICRF), 230
 International Charter on Space and Major Disasters, 135, 264, 334
 International collaboration, 284–286
 International Council for the Exploration of the Seas (ICES), 202
 International Council of Scientific Unions (ICSU), 128
 International Geophysical Year (IGY), 128
 International Satellite Cloud Climatology Project (ISCCP), 810
 International Space Station (ISS), 415
 International Telecommunication Union (ITU), 639
 Intertropical convergence zone (ITCZ), 69, 70
 Intraseasonal oscillation (ISO), 349
 Ionospheric F-region, 361
 Ionospheric propagation, 286–290
 Appleton-Hartree formula, 286, 287
 complex refractive index, 286
 constitutive relations, 286
 incoherent backscatter radar, 289
 ionosonde, 288–289
 magnetoionic theory, 287
 Maxwell's equations, 286
 non-uniform magnetic field, 286
 orthogonal coordinate system, 286, 287
 plane-polarized wave, 286
 plasma frequency, 287
 radio waves, 286
 space-based remote sensing, 289–290
 Iridium satellite constellation, 431–433
 Irrigation management, 291–299
 airborne platforms, 296
 canopy-air temperature difference, 296
 crop water stress index (CWSI), 293–294
 data networks, 298
 energy balance model, 296
 field surface energy balance (FSEB), 293, 295, 297
 low-cost microcontroller, 298
 neutron moisture meter (NMM), 292
 normalized difference vegetative index (NDVI), 293
 on-farm water management, 298
 Penman's equation, 292
 plant characteristics, 291
 plant water stress, 293
 plant water use, 292
 radio frequency technology, 298
 remote sensing, 292–293
 self-propelled irrigation systems, 297
 software design, 297
 soil roughness, 296
 soil water content, 292
 spatial and temporal resolution, 296
 spatial and temporal variability, 292
 time-temperature threshold index (TTTI), 293–295
 water-soluble nutrients, 297
 IsatDataPro message-based service, 430
 IsatM2M message-based service, 430
 Isotropic scattering, 586
 Iterative extended boundary condition method (IEBCM), 157
- J**
 Jakobshavn Glacier in Greenland, 278
 Jason-II advanced microwave radiometer, 187

- K**
 Kanton Island, 349
 Kelvin waves, 351
 Kirchhoff model (KM), 154
 Kirchhoff's law, 403
 Klein-Nishina formula, 220
 Klyuchevskoi volcano, 892
 Kramers-Kronig relations, 363, 569–571
- L**
 La Conchita landslide in California, 328, 329
 Lambertian surface, 617, 626
 Lambert's law, 14–15, 626
 Land-atmosphere interactions, 325–328
 Land cover classification, decision fusion, 141–143
 consensus theory, 142–143
 distribution-free/nonparametric techniques, 142
 Gaussian maximum likelihood classifier, 141
 hypertemporal data, 143
 import vector machines (IVMs), 142
 multisensor data, 142
 multivariate Gaussian model, 142
 stacked vector approach, 141
 support vector machines (SVMs), 142, 143
 very-high-resolution imagery, 142
 Land Remote Sensing Commercialization Act of 1984, 79, 80, 333
 Land Remote Sensing Policy Act of 1992, 80, 334, 521
 Landsat Data Continuity Mission, 79–80
 Landsat satellites, 687
 Landslides, 328–331
 geological hazards, 328
 Interferometric Synthetic Aperture Radar (InSAR), 330–332
 inventory maps, 328–330
 La Conchita landslide in California, 328, 329
 mass movements and slope failure, 328
 rotational landslide, 328, 329
 stereo aerial photographs, 328
 Land surface emissivity (LSE), 303–310
 absolute emissivity, 308
 alpha-derived emissivity (ADE) method, 309
 alpha-residual method, 307–308
 atmospheric and energy-balance models, 304
 classification-based algorithms, 309–310
 decorrelation contrast stretch, 304, 305
 desert landscape, 304
 emissivity bounds method, 306
 graybody emissivity method, 308
 land surface temperature, 304, 305
 model emissivity method, 308
 normalized emissivity method (NEM), 308
 Planck draping method, 307, 308
 Planck's law, 304
 ratio methods, 306–307
 temperature-emissivity separation algorithm (TES), 309
 temperature-independent spectral indices (TISI) method, 308–309
 thermal infrared (TIR) radiation, 304
 two-time, two-channel approach, 305–306
 Land surface models (LSMs), 904
 Land surface phenology (LSP), 886–889
 MOD12Q2 algorithm analyzes, 889
 optical remote sensing, 887–888
 passive microwave sensors, 888
 satellite active (radar) sensors, 888
 validation, 889
 vegetation canopy status, 886
 Land surface roughness, 311–313
 Land surface temperature (LST), 314–319
 atmospheric effects, 315–317
 brightness temperature, T_b , 318
 color temperature, T_c , 318
 directional effects, 317
 emissivity, 315
 energy-balance and ecological studies, 314
 geothermal and volcanic monitoring, 314
 land surface heterogeneity, 317
 model temperature, T_m , 318
 Planck's law, 315
 spectral radiance, 318
 split-window algorithm, 319
 Stefan-Boltzmann law, 315
 TIR images, 314, 315
 Land surface topography, 320–324
 ASTER, 324
 cartosat and prism, 324
 digital elevation model (DEM), 320–324
 natural processes, 320
 optical imaging systems, 321–322
 orthorectification, 320
 physical processes, 320
 radar satellite systems, 323
 shuttle radar topography mission (SRTM), 323–324
 surface and subsurface water flow, 320
 Système Pour l'Observation de la Terre (SPOT) satellites, 324
 topographic map, 320, 321
 vegetation, 320
 Large-aperture reflector antennas, 675–676
 Larsen B ice shelf, 280
 Last Glacial Maximum, 270
 Latent heat of evaporation (LE), 326
 Laurentide ice sheet, 100
 Law of remote sensing, 333–335
 L-band measurements, 785
 Leaf area index, 211
 Leaf chlorophyll content, 26–27
 Legendre polynomials, 609–610
 Lidar In-space Technology Experiment (LITE), 415
 Life-history methods, 641
 Light detection and ranging (LIDAR), 335–337, 415, 837
 active electro-optical, 697–698
 altimetry and mapping systems, 335
 backscatter lidars, 335–336
 differential absorption lidars, 180, 337
 Doppler lidars, 180–181, 336–337
 emerging technology, 177–182
 sea ice, 149
 techniques, 837
 transmitter and receiver, 335
 Lightning, 339–343
 cloud to ground (CG), 339
 electromagnetic spectrum, 340
 EUMETSAT Meteosat Third Generation (MTG), 343
 Fast On-Orbit Recording of Transient Events (FORTE), 341–342
 Geostationary Operational Environmental Satellite (GOES-R), 343
 ground-based lightning location systems, 340
 Lightning Imaging Sensor (LIS), 342–343
 noncontinuous multi-scale physical process, 339
 Optical Transient Detector (OTD), 341, 342
 Lightning Imaging Sensor (LIS), 418
 Limb sounding, atmospheric, 344–347, 695, 836
 infrared limb sounding, 346
 limb radiances and line broadening, 344–345
 LIMS and SAMS observations, 346
 Mars Climate Sounder instrument (MCS), 344
 microwave and infrared, 344
 microwave limb sounding, 345–346
 nadir sounding, 344
 optimal estimation method, 346
 radio occultation, 345
 retrieval/inverse calculation, 346
 solar occultation observations, 345
 stratospheric and mesospheric composition observations, 346
 visible and ultraviolet, 346
 Linearly polarized wave, 579
 Line emission, 581–582
 Livestock grazing, 653
 Living marine resources (LMRs), 202
 Living Planet, 896
 Local thermodynamic equilibrium (LTE), 495, 835
 Long-wave infrared (LWIR) spectrum, 692
 Lorentz process, 568–570
 Low earth orbit (LEO) satellite, 430–433
 ARGOS, 430, 432
 Globalstar, 433
 Iridium, 431–433
 Orbcomm operates, 433
 price comparisons, 433
 weather prediction, 419
 Lunar Orbiter Laser Altimeter (LOLA), 335
 Lunar radiation, 593
- M**
 Madden-Julian Oscillation (MJO), 349–357
 atmospheric and oceanic parameters, 350
 cloudiness, 349
 cold high clouds, 351
 convective active phase, 349
 convective cloudiness, 351
 convectively inactive phase, 349
 convective region, 352
 CPC Merged Analysis of Precipitation (CMAP), 351
 eddy available potential energy (EAPE), 353
 Global Precipitation Climatology Project (GPCP), 352
 impacts, 356–358
 infrared brightness, 351
 infrared window radiance, 350
 intraseasonal oscillation (ISO), 349
 Kelvin waves, 351
 life cycle, 353
 Microwave Sounding Unit (MSU), 351
 moist thermodynamic structure, 353
 moisture structure, 353
 NOAA polar orbiter satellites, 350–351
 Northern Hemisphere, 350
 ocean surface fluxes, 353–354
 outgoing longwave radiation (OLR), 350
 pressure-longitude cross sections, 353, 354
 QuikSCAT surface winds, 352
 rainfall, 349
 rawinsonde and surface station data, 350
 Rossby wave, 352
 satellite remote sensing data, 350
 seasonal-to-interannual climate predictions, 350
 sea surface temperature, 355
 spatial-temporal evolution, 353
 thermodynamic environment and surface conditions, 352
 three-dimensional temperature, 353
 TRMM Multisatellite Precipitation Analysis (TMPA) project, 352
 tropical ocean-atmosphere system, 349
 Tropical Rainfall Measurement Mission (TRMM), 352
 Magnetic field, 358–361
 data sources, 358–359
 Earth's deep interior, 358
 fluid outer core, 358

- geomagnetic field characteristics, 359–360
 ionospheric F-region, 361
 Laplacian potential field, 361
 pre-Maxwell approximation, 360–361
 remote sensing applications, 361
 Magnetic permeability, 571
 Man computer Interactive Data Access System (McIDAS), 849
 Mangrove ecosystem, 74
 Manned aircraft, 409–410
 Marine cryosphere, 92
 Marine Mammal Research Program (MMRP), 10
 Mars Climate Sounder instrument (MCS), 344
 Mass-budget approach, 276
 Mass continuity, 769
 Mauna Loa, Hawaii, 314, 315
 Maxwell's equations, 571, 572, 579
 boundary conditions, 151
 charge density and current density, 150
 electromagnetic scattering, 156
 Foldy-Lax multiple scattering equations, 598
 QCA and NMM3D, 598–600
 scalar and vector potentials, 151
 scattering by dense media, 597–598
 Measurement of Air Pollution from Satellites (MAPS) experiment, 32
 Measurements of Pollution in the Troposphere (MOPITT) instrument, 32
 Media, electromagnetic characteristics, 362–364
 dispersion, 363
 inhomogeneous composite, 363
 macroscopic properties, 362
 metamaterial, 363–364
 Megha-Tropiques ("Clouds-Tropics"), 417
 Mercury cadmium telluride (MCT) detectors, 832
 Meridional heat transport (MHT), 486–487
 Mesocyclone, 772
 Mesoscale convective systems (MCSs), 775
 METOP-SG satellite, 844–845
 Michigan Microwave Canopy Scattering model (MIMICS), 600
 Microbolometers, 832
 Microbursts, 771
 Microelectromechanical systems (MEMS), 186, 679
 Micro Pulse Lidar Network (MPLNET), 18, 336
 Microstrip detectors, 221–222
 Microwave Analysis and Detection of Rain and Atmospheric Structures (MADRAS), 417
 Microwave dielectric properties of materials, 364–373
 dry snow, 368
 electromagnetic quantities, 364–365
 freshwater, 366
 pure and freshwater ice, 370
 sea ice, 370–371
 seawater, 366–367
 soils, 371–373
 vegetation, 373
 water, 366
 wet snow, 369
 Microwave horn antennas, 375–382
 bandwidth, 375–376
 beamwidths, 375
 circular horn, 375, 376
 conical horns, 379, 380
 corrugated horns, 375, 376, 380–381
 dielectric horns, 382
 electromagnetic signals, 375
 functioning environment, 377
 gain and aperture efficiency, 376
 matched horn feeds, 381
 multimode horns, 380
 polarization, 376–377
 primary types, 375
 pyramidal horns, 375, 376
 rectangular horns, 377–380
 Microwave Imaging Radiometer with Aperture Synthesis (MIRAS), 785
 Microwave Limb Sounder (MLS), 356, 911
 Microwave radiometers, 46–47, 382–385
 antenna effects, 46
 antenna temperature, 383
 blackbodies, 383
 brightness temperature, 383
 calibration, 384–385
 conventional radiometers, 386–389
 correlation radiometer, 384, 389–390
 Dicke radiometer, 384
 interferometers, 390–394
 polarimeters, 394–398
 radiated energy, 383
 radiometer antennas, 384
 Rayleigh-Jeans approximation, 382
 receiver effects, 46–47
 sensitivity, 383
 stability and accuracy, 383
 total power radiometer, 383, 384
 Microwave region, 577
 Microwave scatterometers, 533
 Microwave Sounding Unit (MSU), 351
 Microwave subsurface propagation and scattering, 398–402
 airborne sensing (CARABAS, radiometers), 401–402
 attenuation and dispersion, 400–401
 ground-penetrating radar (GPR), 401
 monochromatic electromagnetic waves, 398
 soil, rock, ice, and snow dielectric properties, 399–400
 spaceborne sensing, 402
 subsurface scattering and sensing, 401
 Microwave surface scattering and emission, 403–405
 Middle ultraviolet (MUV) remote sensing, 853–854
 Mid-wave infrared (MWIR) spectrum, 692
 Mie theory, 368
 Milankovitch cycles, 100
 Millimeter-wave spectrum, 692
 Mineral exploration, 705–711
 gold exploration, 705, 709–711
 HyMap data of Mt. Fitton area, 705, 706
 silica, alunite, kaolinite, 705–709
 Minimum–maximum difference (MMD) algorithm, 309
 Minnaert model, 617
 Mirror surface, 617
 Misocyclones, 777
 Mission costs of earth-observing satellites, 405–407
 Mission operations, science applications/requirements, 407–408
 Moderate Resolution Imaging Spectroradiometer (MODIS), 735, 822, 823
 aerosols, 18
 atmospheric effects, 315–317
 Chaiten volcano, 891, 892
 cloud properties, 71
 fire, 215
 glacier mapping, 106
 glaciers, 106
 Madden-Julian Oscillation (MJO), 356
 ocean color, 438
 on-board calibrators, 50
 sea surface temperature, 756
 surface water, 817
 vegetation phenology, 887
 Modified chlorophyll absorption ratio index (MCARI), 26
 Modular Optoelectronic Multispectral Scanner (MOMS–02/D2), 416
 Moist adiabatic lapse rate, 768
 Molecular absorption, 493–495
 electronic energy, 493
 energy associated to rotation, 493
 K-distribution (KD), 494–495
 line absorption/emission, 493
 line broadening, 493–494
 line shape, 494
 Planck constant, 493
 regular and random models, 494
 translational energy, 493
 vibrational mode, 493
 Molniya orbit, 416
 Monochromatic electromagnetic waves, 398
 Monte Carlo methods, 607, 619–622
 1D to 3D Monte Carlo, 620–622
 local estimation, 620
 Russian roulette, 619–620
 Montreal Protocol on Substances that Deplete the Ozone Layer, 799
 Moore's Law, 160
 Moving Ship Tomography, 9
 Mt. Pinatubo eruption in 1991, 17
 Multi-angle Imaging SpectroRadiometer (MISR), 48, 658
 aerosols, 18
 ice albedo, 111
 on-board calibrators, 50
 Multimode horns, 380
 Multiple beams, 668
 Multiple scattering, 585–587
 analytical theory, 586–587
 and radiative transfer, 585
 random distribution of particles, 585–586
 Multi-polarization multidirectional multispectral instrument (3MI), 666
 Multispectral imagery, 693, 694
 Multispectral Thermal Imager (MTI), 659–660
N
 Nadir sounding, 836
 National Elevation Dataset (NED), 320
 National Oceanic and Atmospheric Administration (NOAA), 686
 National Operational Hydrologic Remote Sensing Center (NOHRSC), 823
 National Polar-orbiting Operational Environmental Satellite System (NPOESS), 417–418
 National Research Council (NRC, 2008), 87
 National Satellite Land Remote Sensing Data Archive (NSLRSDA), 135
 National Snow and Ice Data Center (NSIDC), 105
 Natural radiation within Earth's environment, 558–573
 Navier–Stokes equations, 35
 Near Earth Asteroid Rendezvous Shoemaker spacecraft, 227
 Near infrared (NIR) spectrum, 692
 Near real time (NRT) measurement, 843
 Near ultraviolet (NUV) remote sensing, 853–854
 Neural networks, 506–507
 Neutron moisture meter (NMM), 292
 Night-light (NL), 871–872
 Nimbus–7, 686
 NIR aerosol reflectance ratio, 439
 NLSST, 755
 NOAA–6, 686–687
 Noctilucent clouds, 418, 664

- Noise injection radiometer (NIR), 386–387
 Non Local Thermodynamic Equilibrium (NLTE), 495
 Normalized difference vegetation index (NDVI), 293
 crop biomass, 27
 irrigation management, 293
 vegetation, 62
 vegetation indices, 883
 Normalized emissivity method (NEM), 308
 Normalized radar cross section, 533, 534
 North America-Europe data exchange group (NAEDX), 264
 North Atlantic iceberg dataset, 281
 North Atlantic right whale, 206
 Northwestern Hawaiian Islands (NWHI), 206
 Nowcasting Satellite Application Facility (NWC SAF), 811
 Numerical Maxwell model of 3-dimensional (NMM3D) method, 598–600
 Numerical weather prediction (NWP), 912–919
 atmospheric data assimilation, 912–913
 data impact, 916–918
 GMF calibration, 57
 numerical atmospheric modeling, 912
 satellite data products, 913
 satellite data usage, 913–916
 sea surface temperature, 757
 Nyquist-Kotelnikov's theorem, 616
- O**
 OAIS Reference Model, 123–126
 context information, 124–125
 data rescue, 123
 Designated Community, 123, 126
 disaster and security management plans, 125
 documented policies and procedures, 125
 FGDC/ISO metadata standards, 126
 information packages, 124, 125
 preservation description information, 124
 Preservation Metadata: Implementation Strategies (PREMIS), 125
 provenance information, 124
 sufficient control, 123
 XML Formatted Data Unit (XFDU), 125
 Observing System Experiment (OSE), 917
 Ocean acoustic tomography, 4–10
 acoustic signals, 5, 6
 Central Equatorial Pacific Tomography Experiment (CEPTE), 8
 Fram Strait, 7
 Greenland Sea Project, 7
 marine mammals, 9–10
 mesoscale variability, 5
 Moving Ship Tomography, 9
 oceanic sound speed profile, 4
 passive tomography, 9
 rays and modes, 7
 RTE87, 8
 sound speed and current, 5–7
 thermometry, 8–9
 vertical line array (VLA), 9
 Webb Research swept-frequency acoustic source, 9
 Ocean applications of interferometric SAR, 426–428
 Ocean-atmosphere water flux and evaporation, 480–488
 atmospheric circulation, 480, 481
 bulk parameterization, supply side estimation, 481–483
 divergence of moisture transport, demand side estimation, 482–485
 equation of water conservation, 480
 marine atmosphere water conservation, 484–486
 ocean heat and surface salinity conservation, 485–488
 water cycle, 481
 Ocean color, 437–444
 applications, 441–443
 atmospheric correction (AC), 439
 Coastal Zone Color Scanner (CZCS), 437
 chlorophyll-*a* (Chl-*a*), 438–441
 colored dissolved organic matter (CDOM), 438, 440
 datasets, 441
 Medium Resolution Imaging Spectrometer (MERIS), 438
 Moderate-Resolution Imaging Spectroradiometer (MODIS), 438
 Modular Optoelectronic Scanner (MOS), 437
 NIR aerosol reflectance ratio, 439
 normalized specific absorption, 440
 Ocean & Land Colour Instrument (OLCI), 438
 relative contributions, 439
 Sea-viewing Wide Field-of-view Sensor (SeaWiFS), 437–438
 Spaceborne ocean color imagery, 437
 Suspended Particulate Matter (SPM), 438, 440
 total backscattering and absorption, 440
 Visible Infrared Imaging Radiometer Suite (VIIRS), 438
 Ocean data assimilation, 450–453
 applications, 451–453
 atmospheric weather prediction, 450
 constrained least-squares problems, 451
 filters and smoothers, 451, 452
 inverse methods, 451
 model-data misfit, 450–451
 satellite data, 453
 Ocean data telemetry, 429–433
 geostationary satellite, 429–430
 low earth orbit satellite, 430–433
 terrestrial radio, 429
 Ocean internal waves, 433–436
 acoustic Doppler current profiler (ADCP), 434
 Andaman Sea, 434–435
 conductivity-temperature-depth (CTD), 434
 internal solitons/internal solitary waves (ISWs), 434
 Mediterranean Sea, 436
 optical sensors flying detector, 433
 sea surface elevation, 433
 South China Sea, 435–436
 Strait of Gibraltar, 433
 synthetic aperture radar (SAR), 433
 thermocline, 434
 tidal flow, 434
 Ocean modeling, 446–450
 Arakawa C-grid, 448, 449
 climate-oriented sensitivity experiments of predictions, 449
 computational fluid dynamics (CFD), 449
 direct numerical simulations (DNS), 449
 equations of motion, 448
 finite volume elements, 448
 forced-dissipative fluid, 447
 general ocean circulation, 447
 horizontal model grid, 448, 449
 large-scale eddy simulations ("LES"), 449
 near-surface flow and temperature fields, 449, 450
 operational now-and forecasting, 449
 process-oriented modeling, 449
 satellite data, 450
 spatial and temporal scales, 447
 Ocean remote sensing, 469–478
 chlorophyll-*a* concentration, 472–473
 electromagnetic radiation (EMR), 472
 high-altitude platforms, 469
 ocean data products, 474–475
 ocean phenomena, 475–477
 ocean properties measurement, 469
 ocean sampling capability, 471
 operational applications, 478
 passive infrared radiometers, 472
 passive microwave radiometers, 472–473
 polar-orbiting satellites, 469
 scanning radiometer, 469
 scatterometer, 473–474
 Southern Ocean, 470
 space-time sampling capabilities, 471
 visible waveband radiometers, 472
 Ocean salinity, 420
 Ocean surface fluxes, 353–354
 Ocean surface topography (OST), 455–461
 basin-scale variability, 457, 459
 conventional radar altimetry, 460
 density field, 456
 geoid, 456
 geostrophic balance, 455
 mesoscale variability, 459
 ocean currents, 456
 ocean general circulation, 456, 458–459
 practical applications, 459–460
 Rossby number, 455
 SAR interferometry, 461
 satellite altimetry, 455, 457–458
 satellite mission called Surface Water and Ocean Topography (SWOT), 461
 Seasat, 460
 turbulence and white caps, 455
 Ocean surface velocity (OST), 461–468
 Agulhas Current, 466–468
 conventional along-track interferometry techniques, 462
 2D barotropic tidal current model, 462
 Doppler anomaly estimation, 462
 Doppler anomaly signal, 464–465
 mesoscale and submesoscale processes, 461
 ocean basin-scale range doppler velocity monitoring, 466
 range doppler velocity retrievals, 462–464
 satellite altimetry, 461
 spaceborne altimetric range measurements, 462
 SRTM mission, 462
 Offset (symmetric) parabolic reflectors, 669–674
 edge and feed tapers, 669–670
 geometrical parameters, 669
 radiation pattern characteristics, 670–674
 On-board calibrators (OBC), 50
 Onion peeling approach, 346
 On-orbit calibration, 49–51
 cross-calibration, 50–51
 on-board calibrators (OBC), 50
 unattended desert sites, 50
 vicarious calibration (VC), 50
 Open Data Access Protocol (OpenDAP), 121
 OpenPort, 432
 Open-source Project for a Network Data Access Protocol (OpenDAP), 126
 Operational Linescan System (OLS), 871
 Operational transition, 489–492
 Bayh-Dole Act of 1980, 490–491
 characteristics of, 491
 successful and unsuccessful transitions, 491–492
 transition pathways, 490

- Optical/infrared, radiative transfer, 495–501
 atmospheric radiation, 496
 elastic scattering, 499–501
 radiative processes and radiative transfer equation, 496–498
- Optical Transient Detector (OTD), 341, 342
- Orbiting Carbon Observatory (OCO), 418
- Orthorectification, 320
- Orthotropic case, 617
- Outgoing longwave radiation (OLR), 350
- Owens Valley Solar Array (OVSA), 593
- Ozone, 60–61, 839
- Ozone hole, 797
- Ozone profiles, 855
- Ozone recovery and climate change, 799–800
- Ozone Mapping and Profiler Suite (OMPS), 857, 865–867
- Ozone Monitoring Instrument (OMI), 836, 839–841, 863–865
- P**
- Pancake ice, 729
- Panchromatic imagery, 693
- Panfilov model, 617
- Parabolic arc reflector antennas, 670–671
- Parabolic cylinder, 675
- Parabolic-reflector antennas, 669–674
- Parabolic torus reflector, 675
- Park cool island (PCI) effects, 880
- Partially polarized wave, 580
- Passive electro-optical instruments, 693–697
- Passive microwave imaging radiometers, 698–699
- Passive microwave radiation, 921
- Passive microwave sensing, 634–639. *See also* Radio-frequency interference (RFI)
- Passive remote sensing techniques, 574
- Patagonian ice sheet, 100
- Pattern recognition and classification, 503–508
 land cover maps, 503
 remote sensing data, 503–504
 supervised methods, 504–508
 unsupervised techniques, 504
- Peat, 802
- Pencil-beam reflectors, 668
- Penman-Monteith equation, 326
- Perfect Electrical Conductor (PEC), 559
- Permanent Scatterer Technique PSInSARTM, 331
- Perpendicular vegetative index (PVI), 26
- Persistent Scatterer Interferometry (PSInSAR), 804, 805
- Phase velocity, 580
- Photogrammetric Engineering and Remote Sensing in 2003, 24
- Physical geodesy, 230–231
- Physical optics (PO), 668
- Piezoelectric transducers (PZTs), 695
- Pilot Evaluation of High-Resolution Precipitation Products (PEHRPP), 650
- Pine Island Glacier, 279
- Pinty-Dickinson-Verstraete model, 617–618
- Pixel detectors, 222
- Planck draping method, 307, 308
- Planck emission, 574–576
- Planck equation, 754
- Planck radiation, 559–562
 electromagnetic modes, 559, 561
 Perfect Electrical Conductor (PEC), 559
 Rayleigh-Jeans approximation, 561–562
 spectral brightness, 560–561
 standing wave modes, 559–560
- Planck's law, 304, 315
- Plane acoustic wave, 2
- Plant community, 653–654
- Plate tectonics, 147
- Pointing, acquisition, and tracking (PAT), 171–172
- Point matching method (PMM), 157
- Polar ice dynamics, 509–511
 data acquisitions, 510
 feature tracking, 510
 ice cover and displacement, 509
 ice floe interactions, 509
 ice movement, 509
 image correlation, 510
 interferometry, 510
 synthetic aperture radar (SAR), 510
- Polarimeters, 696
- Polarimetric calibration, 697
- Polarimetric lidar, 180
- Polarimetric microwave radiometers, 395–397
 ocean surface wind measurements, 397
 polarization combining radiometer (indirect method), 397
 polarization-correlating radiometer (direct method), 396–397
 Stokes' parameters, 395
 WindSat, 397
- Polarimetric remote sensing, 663–666
 from airplanes, 664–665
 from balloons, 663
 from rockets, 664
 from space, 665–666
- Polarimetric SAR data, 53
- Polarization, 588–590
- POLARization and Directionality of Earth's Reflectance (POLDER), 18, 659
- Polarization diversity, 579–580
- Polarization ratio, 109
- Polar mesospheric clouds, 664
- Polar ocean navigation, 512–514
 ice centers and commercial providers, 514
 ice information, 512–513
 observing systems, 513
 satellite microwave sensors, 512
 satellite remote sensing, 514
 shipboard radars, 512
- Polar Orbiting Environmental Satellites (POES), 686
- Polar region observing system
 albedo and surface temperature measurements, 96
 ground-based observations, 96
 harsh weather conditions, 92
 ice sheets, 94–95
 measurement reference frames, 97
 satellite communications and positioning, 96–97
 satellite elements, 92
 sea ice extent and concentration, 95
 sea ice thickness and dynamics, 95–96
- Polar stratospheric cloud (PSC) particles, 834
- Polder-Van Santen model, 369
- Policies and economics, 515
- Post-Mayo-Krimmel Collection, 104
- Potential fishing zones (PFZ), 204
- Potter horn, 380
- Power flux density, 579
- Poynting's theorem, 152
- Practical Salinity Scale (PSS), 747
- Precipitation, climate monitoring and prediction
 CPC Merged Analysis of Precipitation (CMAP), 61
 Global Precipitation Climatology Project (GPCP), 61
 Precipitation Radar, 61
 rainfall, 61
- Precipitation efficiency, 769
- Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSSIAN), 900
- Precision agriculture, 515–517
 agronomical, 516
 modern crop management, 515
 remote sensing technology, 516
 socioeconomic obstacles, 516
 spatial and temporal variability, 516
 technology obstacles, 516–517
- Precision orbit determination (POD), 526, 529
- Pre-Maxwell approximation, 360–361
- Primary light scattering, 612–613
- Probability distribution function (PDF), 249
- Processing levels, 517–520
 Committee on Data Management and Computation (CODMAC), 518
 Data Processing Levels, 518
 data processing requirements, 519–520
 Global Atmospheric Research Program (GARP), 518
 implementations, 520
 levels of service (LOS), 519
 World Climate Program (WCRP), 518
- Program for Evaluation of High Resolution Precipitation Products (PEHRPP), 901
- Project for On-Board Autonomy (PROBA) technology, 415
- Project Mercury, 681
- Propagation mechanism, 201
- Pseudo-source function, 585
- Public-private partnerships, 520–522
 government organization, 520
 infrastructure development, 521
 LightSAR, 522
 partnerships and remote sensing policy, 521
 railway lines and bridges, 521
 Sea-viewing Wide Field-of-view Sensor (SeaWiFS), 522
 special-purpose vehicle (SPV), 520
 Surrey Satellite Technology Ltd. (SSTL), 522
 synthetic aperture radar (SAR), 522
 US Landsat system, 522
- Pulse-limited altimeters, 530
- Pulse-position modulation (PPM)
 efficiency of, 167
 laser transmitter, 167–168
 receive electronics, 168–169
 serially concatenated codes, 166
 single-photon-sensitive photodetectors, 168
- Pulse repetition frequency (PRF), 139, 542–543
- Pulse severe storm, 771
- Push-broom radiometer system, 388–389
- Pyramidal horns, 375, 376
- Q**
- Quantum electrodynamics, 578
- Quantum radiation, 578–579
- Quantum well infrared photodetectors (QWIPs), 832
- Quasicrystalline approximation (QCA), 597–600
- Quick Scatterometer (QuikSCAT), 63
- QuikSCAT satellite, 419–420, 534, 823–828, 872–874
- R**
- Radar altimeters, 525–531, 556
 brown waveform, 527–528
 flight systems, 528
 ice-sheet radar altimetry, 528–529
 measurement precision, 527, 528
 mesoscale features, 526
 near-polar ice, 527
 oceanographic vs. geodetic orbits, 526–527

- orbit considerations, 529–530
- propagation delay corrections, 527
- pulse-limited geometry, 527, 528, 530
- satellite-based radar altimeter, 530–531
- sea-surface corrections, 527
- sea-surface height (SSH), 526
- significant wave height (SWH), 526
- Skylab, 525
- wind speed (WS), 526
- Radar Imaging Satellite (RISAT), 422
- RADARSAT, 681, 683–684
- RADARSAT–1, 283
- RADARSAT–1 Antarctic Mapping Project (RAMP), 111
- Radar scatterometers, 532–535, 557
 - applications, 533–534
 - backscatter cross section, 533
 - fundamentals, 533
 - microwave scatterometer, 533
 - spaceborne scatterometer, 533–535
 - wind scatterometer, 532, 534
- Radiant intensity, 579
- Radiation, 558–573, 588, 589
 - direct ray, 593
 - at L-band, 583
 - lunar, 593
 - multiple scattering, 585–587
 - nonthermal radiation, 558
 - reflected ray, 594
 - solar, 591–593
 - sources and characteristics, 574–576
 - thermal radiation, 558–573
 - volume scattering, 595–605
- Radiation pattern characteristics, 670
- Radiation sources
 - and characteristics, 574–576
 - continuum radiation, 582–584
 - cosmic microwave background, 581
 - line emission, 581–582
- Radiative Flux Assessment, 810
- Radiative transfer equation (RTE), 564, 585, 626–629
 - differential equation formulation, 628–629
 - volumetric interaction parameterization, 627–628
- Radiative transfer models, 854
- Radiative transfer, solution techniques, 606–622
 - analytical solutions, 607
 - discrete ordinate method, 618–619
 - Fourier series expansions, 609–611
 - Gauss-Seidel method, 613–615
 - initial boundary value problem, 607–609
 - input optical models, 615–618
 - iterative scheme, 611
 - Monte Carlo methods, 607, 619–622
 - numerical solutions, 607
 - primary light scattering, 612–613
 - successive approximation methods, 611–612
- Radiative transfer theory (RTT)
 - history of, 624–625
 - integral formulation and received power, 629, 630
 - Maxwell differential equations, 624
 - multiple scattering, 585–587
 - numerical and approximate solutions, 632–633
 - of partially polarized radiation, 630–632
 - radiative transfer equation (RTE), 626–629
 - specific intensity, 625–627
 - volume scattering, 595–597
 - wave analytical theory (WAT), 623, 633–634
- Radio detection and ranging (RADAR), 547–557
 - accuracy, 551–552
 - altimeters, 556
 - band designations, 548–549
 - coherence, 555
 - multi-channel radars, 555
 - noise factor, 553
 - polarization, 549
 - precision, 552–553
 - processor, 554
 - pulse compression and bandwidth, 554–555
 - quantization noise, 553
 - radar equation, 553, 554
 - reflectivity, 549–551
 - resolution, 549
 - scatterometers, 557
 - Seasat satellite, 548
 - signal-to-noise ratio (SNR), 553
 - Sir Watson-Watt (father of radar), 548
 - sounders, 557
 - space-based radars, 556
 - synthetic aperture radar (SAR), 548, 556
- Radio-frequency interference (RFI), 634–639
 - anomalies, 637, 638
 - frequency allocations and regulation, 638–639
 - level of contamination, 635
 - mitigation techniques, 637–638
 - out-of-band (OOB) signals, 635
- Radiometric calibration, 48–49, 697
 - band-averaged radiances, 48–49
 - standard filament lamps, 49
- Radiometric remote sensing, 186–190
 - back end receives, 189–190
 - calibration and modulation, 188–189
 - coherent detection, 186, 187
 - front-end technologies, 187–188
 - incoherent detectors, 186
 - Jason-II advanced microwave radiometer, 187
- Radiometric sensitivity, 393
- Radio Solar Telescope Network (RSTN), 593
- Radio wave region, 577
- Rainfall, 61, 418, 640–650
- Rainfall active microwave methods, 644–647
 - CloudSat, 646, 647
 - EarthCARE, 646
 - GPM-DPR, 646
 - path-integrated attenuation (PIA), 644
 - surface reference technique (SRT), 644
 - TRMM PR, 644–646
- Rainfall multisensor datasets, 647–650
 - GPCP and CMAP, 647–649
 - high-resolution precipitation products (HRPPs), 649–650
 - WetNet, 647
- Rainfall passive microwave methods, 642–644
 - emission-based methods, 643
 - ice particles, 642
 - precipitation-size drops, 642
 - scattering-based methods, 643–645
- Rainfall satellite measurements and attributes, 640
- Rainfall visible and infrared methods, 640–642
 - bi-spectral methods, 641
 - cloud indexing methods, 640–641
 - cloud model techniques, 641
 - life-history methods, 641
 - NWP-adjusted and multispectral methods, 641–642
- Raman lidars, 336
- Range-Doppler radar, 556
- Rangeland health, 655
- Rangelands and grazing, 653–656
 - ecological sites, 653–655
 - ecosystem degradation, 655
 - invasive weeds detection, 655
 - livestock grazing, 653
 - rangeland productivity, 655
- risk management with AVHRR NDVI, 656
- very large-scale aerial (VLSA) photography, 655–656
- RapidEye constellation, 423
- Rarefied distribution, 585, 586
- Rayleigh-Jeans approximation, 382, 561–562, 564, 575
- Rayleigh scattering, 854
- Receiver effects, 46–47
- Receiver noise temperature (Trec), 47
- Reciprocal tomography (RTE87), 6
- Rectangular horns, 377–380
- Reflected ray, 594
- Reflected solar radiation sensors
 - multiangle imaging, 658–661
 - polarimetric remote sensing, 663–666
- Reflector antennas, 668–679
 - diffraction analysis techniques, 668–669
 - folded optics, 676–679
 - large-aperture reflector antennas, 675–676
 - offset (symmetric) parabolic reflectors, 669–674
 - parabolic arc, 670–671
 - parabolic cylinder, 675
 - parabolic torus, 675
 - pencil-beam reflectors, 668
 - spherical reflector antennas, 675
- Regional Hydroclimate Projects, 808
- Relaxation frequency, 366
- Remnant magnetization, 232, 359
- Remote sensing
 - active electro-optical instruments, 697–698
 - active microwave instruments, 699–701
 - aircraft, 692
 - geologic structure, 681–684
 - historical perspective, 684–690
 - passive electro-optical instruments, 693–697
 - passive microwave imaging radiometers, 698–699
 - satellite orbits, 691–692
 - spectrum, 692–693
- Remote sensing aircraft, 410–411
- Remote Sensing Space Systems Act of 2005, 81
- Remote Sensing Space Systems Act of 2007, 334
- Repeat-pass interferometry, 426
- Repeat-pass observations, 53
- Republic of China Satellite–1 (ROCSAT–1), 420
- Republic of China Satellite–3/Constellation Observing System for Meteorology (ROCSAT–3/COSMIC), 419
- Research Scanning Polarimeter (RSP), 664
- Resonance fluorescence techniques, 179
- Resource exploration, 702–715
 - direct vs. indirect methods, 702–703
 - gold exploration, 705, 709–711
 - HyMap data of Mt. Fitton area, 705, 706
 - mineral exploration, 705–711
 - multispectral vs. hyperspectral technology, 703–705
 - oil and gas exploration, 710–714
 - silica, alunite, kaolinite, 705–709
- Reststrahlen bands, 304, 308
- Reuven Ramaty High Energy Solar Spectroscopic Imager, 226
- Rice crop scattering model, 604, 605
- Ripple–2 system, 194–195
- Root zone soil moisture, 783, 786
- Ross and Filchner ice shelves, 112
- Rossby number, 455
- Rossby wave dynamics, 352
- Ross Ice Shelf, 272
- Roughness amplitude, 313
- Royal Canadian Air Force (RCAF), 108
- Russian Balkan–1 lidar, 415
- Russian/Finnish Arktika mission, 416

- S**
- Salinity events and thermohaline convection, 750, 751
- Salinity remote sensing, 748–750
- Salinity resolution, 750–752
- San Andreas Fault in California, 148, 683
- SAR-based bathymetry, 719–722
- acoustic sounders, 719
 - bathymetric maps, 721–722
 - radar imaging, 719–720
 - roughness variations, 719
 - underwater topographic features, 720–721
- SAR speckle, 539
- Satélite de Aplicaciones Científicas-C (SAC-C), 419
- Satellite-based radar altimeter, 530–531
- 2008 Satellite Data Security Law, 81
- Satellite observational system, 412–424
- Australian hyperspectral, 414
 - cloud characteristics and aerosols, 418
 - commercial agriculture, 423
 - commercial and hybrid government/commercial high-resolution, 422–423
 - commercial high-resolution remote sensing satellite, 414
 - Compact High-Resolution Imaging Spectrometer (CHRIS), 415
 - cryosphere, 424
 - disaster management, 422
 - Earth Observing System (EOS) Aura satellite, 415
 - Earth radiation budget, 418
 - earth resources and environmental monitoring, 420–422
 - environmental satellites, 413
 - GPS occultation and reflectometry, 419
 - high-resolution SAR, 423
 - hybrid government/commercial systems, 414
 - Hyperion instrument, 415
 - hyperspectral imagers, 415
 - hyperspectral systems, 423
 - lidar, 415
 - lightning, 418
 - microsatellite (microsat) category, 414
 - mission operations, 417
 - multispectral instruments, 415
 - Nimbus-7, 413
 - observational platforms, 415
 - ocean, 419–421
 - rainfall, 418
 - satellite orbits, 415–417
 - soil moisture and freeze/thaw state, 422
 - solid earth, 423–424
 - spacecraft subsystems, 417
 - Space Information-2 (SPIN-2) program, 414
 - stratospheric ozone, greenhouse gases, and trace gases, 418
 - TIROS I, 413
 - tropospheric winds, 418
 - UoSAT-5, 414
 - weather prediction and severe storms, 419
- Satellite orbits, 415–417
- Satellite remote sensors, 870–871
- Satellite sensors, 659–660
- Satellite Supplement, 257
- Satellite-to-satellite tracking, 792–793
- Scanning imaging absorption spectrometer for atmospheric chartography (SCIAMACHY), 665, 840, 842, 856, 863
- Scanning Multichannel Microwave Radiometer (SMMR), 63, 108, 634
- Scatter darkening, 573
- Scattering-based methods, 643–645
- Scattering loss, 573
- Scatterometer Climate Record Pathfinder (SCP) project, 113
- Scatterometers, 700, 872–874
- Scatterometers calibration
- backscatter calibration, 55–56
 - geophysical model function (GMF), 55–57
 - geophysical properties, 55
 - radar backscatter cross section, 54
- Scientific Assessment of Ozone Depletion, 61
- SCISAT-1, 418
- Sea ice, 149
- Sea ice albedo, 722–726
- bidirectional reflectance distribution function (BRDF), 722–723
 - feedback mechanism, 725
 - field portable spectroradiometer, 722, 723
 - impact of surface conditions, 723–726
 - incident solar irradiance, 722
- Sea ice concentration and extent, 727–742
- AMSR-E image, 736, 737
 - Antarctic sea ice, 728
 - Arctic region, 727
 - Beaufort gyre, 732
 - Bootstrap Algorithm, 730–731
 - brightness temperature (TB), 730
 - deep ocean convection, 727
 - ice-albedo feedback effects, 727
 - Moderate Resolution Imaging Spectroradiometer (MODIS) data, 735
 - monthly anomalies and trends, 739–741
 - Northern Hemisphere, 727, 732, 739
 - NT2 Algorithms, 729
 - passive microwave sensors, 729
 - physical and radiative characteristics, 728–729
 - polar heat sink, 727
 - Southern Hemisphere, 733, 735, 739–740
 - surface parameters, 730
- Sea level rise, 743–746
- Argo Project, 744
 - Holocene, 743–744
 - human induced warming, 743
 - paleoclimate evidence, 746
 - satellite era, 745
 - socioeconomic importance, 745
 - thermal expansion, 744
 - tide gauge record, 744
- Seasat SAR, 538
- Seasat-1 Scatterometer (SASS), 534, 535
- Sea surface salinity, 747–753
- brightness temperature (T_B), 747–749
 - contour map, 747, 748
 - Debye equation, 747
 - dissolved inorganic salts, 747
 - error budget analysis, Aquarius/SAC-D satellite, 752, 753
 - Fresnel reflection, 747–748
 - global hydrologic budgets, 750
 - microwave optical depth, 748
 - oceanographic observations dating, 747
 - ocean rainfall estimates, 750
 - Practical Salinity Scale (PSS), 747
 - radiometer antennas, 748
 - salinity events and thermohaline convection, 750, 751
 - salinity remote sensing, 748–750
 - salinity resolution, 750–752
 - science requirements, 750
 - seasonal-interannual climate predictions, 750
 - thermohaline circulation, 747
- Sea surface temperature (SST), 754–758
- accuracy of, 757–758
 - atmospheric transmissivity, 755
 - climate change monitoring, 63
 - diurnal heating, 754
- Earth's climate system, 757
- infrared and radiometers, 755–757
 - microwave radiometers, 757
 - numerical weather prediction (NWP) models, 757
 - Planck equation, 754
 - polar-orbiting satellites, 755
 - radiometers, 754
 - skin effect, 754
- Sea surface wind/stress vector, 759–766
- bulk transfer coefficients, 760, 761
 - direct stress retrieval, 765
 - drag coefficients, 760, 761
 - flux-profile relation, 761, 762
 - microwave radiometer, 765, 766
 - microwave scatterometer, 760
 - numerical weather prediction (NWP), 759
 - ocean current, 760
 - polar-orbiting scatterometer, 766
 - primary functions, 760
 - scatterometry, 761–765
 - space-based microwave sensors, 759
 - stability and surface molecular constraints, 761
 - surface wind convergence, 760
 - synthetic aperture radar (SAR), 766
 - transfer coefficient, 760
- Sea turtles, 207
- Sea-viewing Wide Field-of-view Sensor (SeaWiFS), 80
- fishers, 204
 - Madden-Julian Oscillation (MJO) impacts, 356
 - ocean color, 420, 437
 - public-private partnerships, 522
- SeaWinds scatterometer, 534, 535, 823
- Second-generation GOME (GOME-2) instrument, 865
- Sensor webs, 191–195
- data collector node, 195
 - geophysical products, 191
 - global architecture, 191–192
 - non-rechargeable batteries, 195
 - open-loop approach, 193
 - Ripple-2 system, 194–195
 - satellite retrievals, 191
 - Soil Moisture Active and Passive (SMAP), 422, 193
 - soil moisture fields, 192
 - solar panels and rechargeable batteries, 193
- Service-Oriented Architecture (SOA), 121
- Severe storms, 767–779
- convective storms, 769–773, 778
 - deep convective storm, 768–770
 - hailstone, 772–775
 - strong winds, 775–776
 - tornadoes, 776–778
- Shannon information content, 248
- Short Burst Data (SBD), 432
- Short-wave infrared (SWIR) spectrum, 692
- Shuttle radar topography mission (SRTM), 323–324, 540
- Dutch Wadden Sea, 427
 - Elbe river (Germany), 427
 - land surface topography, 323–324
 - TerraSAR-X Stripmap/Aperture Switching Mode data, 427
- Side-Looking Airborne Radar (SLAR), 108, 113
- Single channel receiver, 386, 387
- Single-look complex (SLC) data, 539
- Single scattering approximation, 586
- Single-scattering optical properties (SSOP), 499–501
- Sinkholes, 802
- Skylab, 525
- Sky irradiance, 309, 316

- Slope statistical Kirchhoff (SSK) model, 154
 Snell-Descartes-Huygens law, 12
 Snell-Descartes law, 15
 Snell's law, 12–13
 Snow accumulation onset date (T_s), 827
 Snowfall, 780–782
 atmospheric temperature profile, 780
 cloud microphysical processes, 780
 frozen vs. liquid precipitation, 781
 snowfall intensity, 781–782
 Snowmelt pattern, 824–827
 Snow water equivalent (SWE), 106, 821
 Sobolev's polynomials, 611
 Soil-adjusted vegetative index (SAVI), 26
 Soil moisture, 783–787
 atmosphere and soil water transfers, 783
 caveats, 786
 disaggregation, 786
 extreme event prediction, 783–784
 field capacity, 783
 L-band measurements, 785
 Microwave Imaging Radiometer with Aperture Synthesis (MIRAS), 785
 microwaves, 785
 remote sensing, 784–785
 root zone, 783, 786
 Soil Moisture Active and Passive (SMAP), 422
 Soil Moisture and Ocean Salinity (SMOS) mission, 785
 soil samples, 784
 spatial resolution, 786
 surface soil moisture, 783
 water-holding capacities, 783
 water resources management, 784
 weather forecast, 783–784
 wilting point, 783
 Soil Moisture Active and Passive (SMAP), 191, 193, 422
 Soil Moisture and Ocean Salinity (SMOS), 907
 ocean salinity, 420
 soil moisture, 785
 Soil properties, 788–790
 optical properties, 788–789
 soil carbon, 789–790
 soil degradation, 789
 space-and airborne imagery, 789
 Soil water content, 83–86
 Solar backscatter ultraviolet (SBUV) instrument, 861, 862
 Solar constant, 806
 Solar microwave bursts (SMB), 593
 Solar radiation, 60, 591–593
 Solid earth mass transport, 791–795
 glacial isostatic adjustment (GIA), 791
 mantle deformation, 793–795
 passive satellites, 791–792
 satellite-to-satellite tracking, 792–793
 Sound amplitude, 1
 Sound Navigation and Ranging (SONAR), 1
 South Falls, 720
 Spaceborne radar topography mission (SRTM), 53
 Spaceborne scatterometers, 533–535
 Spaceborne sensors, 831–832
 Spacecraft subsystems, 417
 Space Information-2 (SPIN-2) program, 414
 Sparse distribution, 585, 586
 Spatially Enhanced Broadband Array Spectrograph System (SEBASS), 306, 308
 Special-purpose vehicle (SPV), 520
 Special Sensor Microwave/Imager (SSM/I)
 cloud liquid water, 68
 sea ice concentration, 63
 Spectral calibration, 697
 Spectral mixing (SM), 872, 873
 Spectral response function (SRF), 48
 Spectral width/bandwidth, 579
 Spectrometers, 695–696, 861
 Specular backscatter, 550
 Spherical divergence, 11
 Spherical harmonics discrete ordinate method (SHDOM), 620
 Spherical reflector antennas, 675
 Split-window algorithm, 319
 Split-window technique, 71–72
 Spot Image, 79
 SpotSAR mode, 544
 Stefan-Boltzmann constant, 807
 Stefan-Boltzmann law, 315, 574–575
 Stefan-Boltzmann relationship, 23
 Stereo imagers, 696
 Stokes intensity vector, 580
 Stokes parameters, 395, 541, 590
 Stokes vector, 663, 665
 Stratospheric Aerosol and Gas Experiment (SAGE) instruments, 17
 Stratospheric ozone, 60–61, 796–800
 chlorine-catalyzed ozone depletion, 797–799
 midlatitude ozone depletion, 799
 ozone hole, 797
 ozone recovery and climate change, 799–800
 production and distribution, 797
 solar ultraviolet radiation, 796
 Strike-slip earthquakes, 148
 Strong fluctuation theory, 368
 Strong wind, 775–776
 Structural geology, 681–684
 digital tectonic activity map of earth, 681, 682
 Project Mercury, 681
 San Andreas fault, California, 683
 sudbury impact structure, Ontario, Canada, 683–684
 Submerged aquatic vegetation (SAV) mapping, 74, 77, 78
 Submillimeter-wave spectrum, 692
 Subsidence, 800–805
 airborne measurement techniques, 804–805
 anthropogenic causes, 802–803
 ground-based measurement techniques, 805
 ground motion, 801
 isostatic movement, 801
 mass movement, 802
 natural causes, 801–802
 seismic activity, 801
 soil consolidation, 802
 soil fabric collapse, 802
 PSInSAR, 805
 remote sensing platform instrumentation, 800
 spaceborne measurement techniques, 803–804
 thermokarst, 802
 volcanic activity, 801–802
 Sudbury impact structure, Ontario, Canada, 683–684
 Sulfur dioxide and volcanic eruptions, 855
 Sun-synchronous orbit, 529–530
 Supercell storm, 771
 Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES) instrument, 837
 Superconductor-insulator-superconductor (SIS), 188
 Supervised methods, 505–508
 decision trees (DT), 505–506
 maximum likelihood classification, 505
 neural networks, 506–507
 support vector machines (SVM), 507–508
 Support vector machines (SVM), 142, 507–508
 Surface mass balance, 269
 Surface Radiation Budget (SRB), 808, 812
 Surface radiative fluxes, 806–812
 Advanced Baseline Imager (ABI), 809
 Advanced Very High Resolution Radiometer (AVHRR), 808
 Afternoon Constellation or A-Train, 809
 albedo, 806
 climate research applications, 807
 cloud-radiative forcing, 807
 direct solar and diffuse sky radiation, 806
 Earth-Sun distance, 806
 geostationary and polar orbiting, 807
 Geostationary Earth Radiation Budget (GERB), 808
 Geostationary Operational Environmental Satellites (GOES), 808
 geostationary satellites, 810–811
 Intergovernmental Panel on Climate Change (IPCC), 808
 International Satellite Cloud Climatology Project (ISCCP), 810
 longwave radiation, 810
 net longwave radiation, 807
 net solar radiation, 806
 Nowcasting Satellite Application Facility (NWC SAF), 811
 Regional Hydroclimate Projects, 808
 shortwave radiation, 809–810
 solar constant, 806
 solar zenith angle, 806
 Stefan-Boltzmann constant, 807
 surface measurements, 811–812
 Surface Radiation Budget (SRB), 808, 812
 temporal and spatial scales, 807
 World Meteorological Organization (WMO), 807
 Surface reflectance matrix, 632
 Surface truth, 815–816
 Surface water, 816–818
 Surface Water Ocean Topography (SWOT), 461, 907
 Surrey Satellite Technology Ltd. (SSTL), 414, 522
 Suspended Particulate Matter (SPM), 438, 440
 Swedish National Space Board (SNSB), 418
 Synthetic aperture radar (SAR), 536–546, 548, 556, 700, 766
 calibration, 51–53
 external SAR calibration, 52
 data processing, 136–140
 fundamental principles, 544–546
 icebergs, 283
 imagery, 536–538
 interferometry, 539–541
 logical functions, 536
 ocean internal waves, 433
 phase, 537
 physical rules, 543–545
 polar ice dynamics, 510
 polarimetry, 541–542
 processor, 537–539
 public-private partnerships, 522
 Système Pour l'Observation de la Terre (SPOT) satellite
 digital elevation model (DEM), 324
 public-private partnerships, 522
T
 TanDEM-X satellite, 423
 Taylor series expansion, 139
 Television Infrared Observation Satellite (TIROS-I), 413
 Television Infrared Observation Satellite-Next Generation (TIROS-N), 350, 686

- Temperature–emissivity separation algorithm (TES), 309
 Temperature-independent spectral indices (TISI) method, 307
 TerraSAR-X Stripmap/Aperture Switching Mode data, 427
 Terrestrial cryosphere, 92
 Terrestrial gamma-ray flashes (TGFs), 226
 Terrestrial radio, 429
 Terrestrial snow, 821–829
 anomalous melt events and ice layers, 826–827
 blended products, 828
 cover significance, 821–822
 depth and water equivalent, 822–826
 disappearance date (T_F), 827
 melt onset date (T_M), 827
 seasonal snow extent, 822, 823
 season length (D_S), 827
 snow accumulation onset date (T_S), 827
 snowmelt pattern, 824–827
 snow water equivalent (SWE), 821
 Terrestrial water storage (TWS), 901, 907
 Thematic climate data records (TCDR), 58
 Thematic Realtime Environmental Distributed Data Services (THREDDS) system, 121
 Thermal inertia, 878
 Thermal radiation, 558–559
 atmospheric opacity, 558
 electromagnetic constitutive properties, 565–571
 emissive power and absorption, 564–565
 emissive power and relative permittivity, 571–573
 invariant brightness temperature, 562–564
 Planck radiation, 559–562
 thermal infrared (TIR) techniques, 558
 thermal microwave techniques, 558
 Thermal radiation sensors (emitted), 830–833
 airborne and spaceborne sensors, 831–832
 calibration and validation, 832–833
 in situ measurements, 831
 Thermocline, 434
 Thermohaline circulation, 747
 Thunderstorm Project, 769
 Thuraya satellite, 430
 Tide gauge record, 744
 Time-temperature threshold index (TTTI), 293–295
 TIROS Operational Vertical Sounder (TOVS) series, 910
 Titan satellite sounding, 2
 TOPEX/Poseidon, 528–531
 Top-of-the-atmosphere radiation (TOA), 145–146
 Topographic map, 320, 321
 Tornadoes, 776–778
 Total internal reflection, 152
 Total mass balance, 269
 Total ozone, 855–857
 Total Ozone Mapping Spectrometer (TOMS), 17, 196, 855–856, 861–863
 Total power radiometer, 386, 387
 Total solar irradiance (TSI), 60
 Trace gases, stratosphere and mesosphere, 834–837
 direct solar absorption, 835–836
 Doppler broadening of spectral lines, 835
 LIDAR, 837
 limb sounding instruments, 836
 local thermodynamic equilibrium (LTE), 835
 nadir sounding instruments, 836
 pressure broadening of spectral lines, 835
 remote sounding observations, 837
 Superconducting Submillimeter-Wave Limb-Emission Sounder (SMILES) instrument, 837
 thermal emission, 834, 836–837
 UV/visible scattering measurements, 836
 zenith observations, 836
 Trace gases, troposphere, 838–845
 detection history from space, 839
 METOP-SG satellite, 844–845
 non-Sunsynchronous orbit, 844
 ozone (O_3), 839
 research themes and operational applications, 841–844
 retrievals and validation, 839–841
 satellite observations, 840–843
 Sun-synchronous orbit, 844
 TROPOMI instrument, 844
 Trafficability of desert terrains, 846–848
 alluvial fans, 847
 badlands/malpais, 847–848
 dry lakes, 847
 dunes/dune fields, 848
 soil composition and moisture, 847
 soil grain size, 847, 848
 subsurface moisture, 848
 surface roughness, 848
 Transantarctic Mountains, 270
 1994 Transarctic Acoustic Propagation (TAP), 8
 Transition zone chlorophyll front (TZCF), 206, 208
 Transmission and reflection, 495
 Transport theory. *See* Radiative transfer theory (RTT)
 Transverse Ranges, 148
 Travel-time sensitivity kernels, 7
 Tropical Rainfall Measurement Mission Precipitation Radar (TRMM PR), 644–646
 Tropical Rainfall Measuring Mission (TRMM) Madden-Julian Oscillation (MJO), 352
 rainfall, 418
 Tropical tropopause layer (TTL), 356
 TROPOMI instrument, 844
 Tropospheric Emission Spectrometer (TES), 33
 Tropospheric ozone, 33, 61
 Tropospheric winds, 849–850
 Transformed chlorophyll absorption ratio index (TCARI), 26
 Transformed soil-adjusted vegetation index (TSAVI), 26
 TurtleWatch, 207
 Twersky-Foldy integral equations, 587, 633
 Twersky theory, 587
U
 Ultraviolet region, 577
 Ultraviolet (UV) remote sensing, 853–859
 airglow, 858–859
 aurora, 859
 from satellites, 854–857
 global space weather systems, 857, 859
 near and middle UV remote sensing, 853–854
 from satellites, 854–857
 second-generation satellite UV, 856–858
 from surface, balloons, and rockets, 854
 Ultraviolet (UV) sensors, 860–868
 backscatter ultraviolet (BUV), 861, 862
 far and extreme UV sensors, 867
 Global Ozone Monitoring Instrument (GOME), 862–863
 ground-based remote sensing, 861
 Ozone Mapping and Profiler Suite (OMPS), 865–867
 Ozone Monitoring Instrument (OMI), 863–865
 rockets, 861
 satellite UV instruments, 861
 Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY), 863
 second-generation GOME (GOME-2) instrument, 865
 solar backscatter ultraviolet (SBUV), 861, 862
 spectrometers, 861
 Total Ozone Mapping Spectrometer (TOMS), 861–863
 Ultraviolet (UV) spectrum, 692
 U.N. Committee on the Peaceful Uses of Outer Space (UNCOPUOS), 332
 Underwater topographic features, 720–721
 Unifying method, 153
 Uninhabited airborne vehicles (UAVs), 415
 UNISPACE III conference in Vienna in 1999, 135
 United Nations Environment Programme (UNEP), 61
 Unmanned aerial vehicles (UAVs), 410
 Unpolarized wave, 580
 Unsupervised techniques, 504
 UoSAT-5, 414
 Upward looking sonar (ULS), 110
 Urban environments, Beijing case study, 869–876
 remote sensing approaches, 869–875
 urban landscape, 869–870
 Urban heat island (UHI), 878–880
 heat stress and health, 878, 880
 land use and, 879–880
 surface temperature, 878
 thermal inertia, 878
 urban materials, 878–879
 US Coast Guard International Ice Patrol (IIP), 281
 US Landsat system, 333
 US Policy on Foreign Access to Remote Sensing Space Capabilities (PDD-23), 80
 UV/visible scattering measurements, 836
V
 Van Cittert-Zernike theorem, 392
 Vatnajökull ice cap, 794
 Vector radiative transfer theory, 601–603
 first-order solution, 602–603
 leaves, branches, and trunks, 601
 phase matrices, 602
 second-order solution, 603
 Vegetation indices, 883–885
 ecosystem processes, 884–885
 leaf area index (LAI), 884
 moderate and coarse resolution, 885
 near-infrared (NIR), 883
 normalized difference vegetation index (NDVI), 883–884
 satellite-based vegetation indices, 883
 soil adjusted vegetation index (SAVI), 884
 vegetation canopy greenness, 883
 Vegetation phenology, 886–889
 environmental conditions, 886
 land atmosphere energy, 886
 land surface phenology (LSP), 886–889
 microwave remote sensing, 888
 optical remote sensing, 887
 water and trace gas exchanges, 886
 Vegetation water content (VWC), 155
 Vegetative indices (VIs), 23–26
 VELMAP project, 111
 Very small aperture terminal (VSAT), 430
 Vertical ozone distribution, 854–855
 Vessel monitoring system (VMS), 206
 Vicarious calibration (VC), 50
 Visible Infrared Imaging Radiometer Suite (VIIRS), 438

- Visible region, 577
 Visible (VIS) spectrum, 692
 Volcanism, 148, 890–893
 deformation detection, 892
 eruption cloud detection, 890–891
 thermal phenomena, 891–892
 Volume scattering, 534, 595–605
 bistatic scattering, 595
 by dense media, 597–598
 coherent superposition, 603, 604
 empirical/simplified physical model, 600–601
 Foldy-Lax multiple scattering equations, 598
 full-wave multiple scattering, 603–605
 incoherent wave, 595
 independent scattering, 597
 QCA and NMM3D results, 598–601
 radiative transfer theory, 595–597
 vector radiative transfer theory, 601–603
 Volumetric absorption coefficient, 627
 Volumetric scattering coefficient, 628
 Vorticity, 776
- W**
 Ward Hunt Ice Shelf, 112
 Water and energy cycles, 895–901
 Blue Planet and Living Planet, 896
 Earth system and water, 895–896
 global hydrological cycle, 899–901
 water flux, 896
 water reserves, 896–897
 Water cycle, 896–899
 Water flux, 896
 Water reserves, 896–897
 Water resources, 903–907
 crop growth, 904–906
 fossil water, 903
 global distribution, 904, 905
 land surface models (LSMs), 904
 remote sensing applications, 906–907
 water balance and water resources, 903–904
 Water vapor, 909–911
 climate feedbacks, 909–910
 clouds and precipitation, 909
 greenhouse gas, 909
 radiative effects, 909–910
 remote sensing, 910–911
 weather influence, 909
 Wave analytical theory (WAT), 624, 633–634
 Wave interference, 580
 Wave propagation, 151–152
 Weather prediction, 912–919
 atmospheric data assimilation, 912–913
 data impact, 916–918
 numerical atmospheric modeling, 912
 satellite data products, 913
 satellite data usage, 913–916
 Weichselian ice sheet, 100
 West Antarctica ice sheet (WAIS), 272
 Wet growth, 773
 Wetlands, 921–922
 WetNet, 647
 Window bands, 692
 Wind scatterometers, 532, 534
 Working Group on Calibration and Validation of CEOS, 40
 Working party 7C (WP7C), 639
 World Data Center (WDC), 104, 128
 World Data System (WDS), 128
 World meteorological organization pace programme (WMOSP), 264
 World Meteorological Organization's (WMO's), 128
 World Radio Conference (WRC), 639
- X**
 Xinchuan Gang Shoal off, 721
 XML Formatted Data Unit (XFDU), 125
 XpressLink, 430
 X-ray region, 577